Beyond Gross National Product: An Exploratory Study of the Relationship between Program for International Student Assessment Scores and Well-being Indices

Chong Ho Yu[1]

[1] Department of Psychology, Azusa Pacific University, California, USA

Correspondence: Chong Ho Yu, Department of Psychology, Azusa Pacific University, 901 East Alosta Avenue, Azusa, California 91702, USA. Tel: 1-626-815-6000. ext. 2749. E-mail: cyu@apu.edu

Abstract

Because national intelligence is a crucial predictor of national wealth, the potential for international test scores to serve as a proxy measure of national cognitive capabilities has intrigued educational researchers. However, national wealth indexes, such as GDP or Purchasing Power Parity (PPP), are of limited use as indicators of quality of life. Since the late 20th and early 21st century, Europeans have been taking the lead in the movement of going beyond GDP by taking social and environmental indicators into account while measuring national progress. Guided by European principles, this study utilizes two indices, Human Development Index (HDI) and Happy Planet Index (HPI), to investigate the extent to which international test scores might predict national well-being. This exploratory study, utilizing correlation matrix, cluster analysis, and conditional regression, found that Program for International Student Assessment (PISA) scores collected in the year 2000 are a significant predictor of national well-being measured by HDI and HPI approximately a decade later. However, when cultural and geopolitical factors were taken into account, it was found that this relationship was not consistent across different subcomponents of HDI and HPI, and across different regions of the globe.

Keywords: European Union Commission, PISA, GDP, well-being, happy planet index, human development index

1. Introduction

1.1 Links between Assessment Scores and National Performance

International test scores, especially in math and science, have been drawing attention from educational researchers since their introduction. This is in part because they are perceived as a proxy measure of national intelligence, which has been considered a crucial predictor of national wealth (Rindermann, 2008). For example, Lynn and Mikk (2007) found that the Trends in International Mathematics and Science Studies (TIMSS) test scores for 8th graders and GDP per capita had a moderate correlation (0.55) whereas national IQ estimates and GDP per capita were correlated at 0.62.

Due to the fact that IQ tests are concerned with abstract reasoning instead of domain-specific knowledge, and transnational IQ estimates are hard to obtain, scholars and politicians turned their attention to the causal link between international math/science assessment and national economic performance. For instance, Hunt and Wittmann (2008) found that the Program for International Student Assessment (PISA) test scores provides an even better prediction of national wealth than TIMSS. Similarly, based on international test scores, including PISA, Hanushek and Woessmann (2008) concluded that the cognitive skills of the population are strongly associated with individual earnings and economic growth. After using the 1960 GDP per capita as the baseline and controlling for the average years of schooling in the same year, increasing test scores by one standard deviation is correlated with increasing the annual growth rates by almost two percent. Following this formula, if the performance of American students in math and science tests were raised to the top level, it would have increased the US growth rate by over a full percentage point. The link between cognitive skills measured by international tests and economic growth was also supported by Organization for Economic Cooperation and Development (OECD), the originator of PISA. According to OECD (2010b), if all 30 members of the OECD raise their average scores on PISA by 25 points, in the next 20 years it would lead to an aggregate gain of $115

trillion in GDP over the lifetime of the generation born in 2010. Although this correlation does not necessarily imply causation, the practical implication for the US is that this modest improvement in PISA could result in $41 trillion gain in US's GDP over 80 years. This prediction is appealing to the US policy makers at the present time of economic recession. It is no wonder many research endeavors and political discussions on PISA tend to focus on GDP growth and other economic indices. For example, after viewing and being impressed by the PISA results, President Obama and Secretary Duncan emphasized the causal link between education and national economic competitiveness while advocating for the American Recovery and Reinvestment Act (West, 2011). However, in the perspective of achieving a well-balanced national-welling, which will be discussed in the next section, it is questionable whether the emphasis on economic output should be the major direction for research and policy-making.

1.2 Purpose

The focus of this article is not to examine concurrent validity between international tests and IQ estimates. Rather, following the line of reasoning that international test scores might indicate national cognitive skills to some degree, the author of this article questions whether using national wealth as the primary, or even the sole, outcome is too narrow to capture the meaning of good science education or good education in general. Classical economics assume human behaviors are driven by rational choices (Bicchieri, 1993). It is reasonable to suppose that a society with an informed public and an engaged, accomplished scientific community should be capable of making wise choices about time allocation and resources used in a number of areas, rather than formulating educational policies with the national goal of single-mindedly boosting competitiveness and GDP. Thus, this article examines the relationship between alternate measures of national well-being, including Human Development Index (HDI) and Happy Planet Index (HPI), and international test scores.

2. Literature Review

In this section the reasons why GDP alone is inadequate as a well-being measure and what remedies have been proposed are discussed. It is important to point out that while the following discussion seems to focus on the European endeavors, the author by no means downplays the contribution to this subject matter by Americans and others. However, at the time of this writing, the literature suggests that Europeans have been taking the lead in this domain. For example, using the keyword "well-being," "wellbeing" or "quality of life," the search for well-being scales or indicators in the US Mental Measurement Yearbook (Buros Institute of Mental Measurements, 2012) returned only 30 results, but none of them addresses different aspects of well-being in a comprehensive fashion, in comparison to what can be seen in the European or international scales. One may argue that Mental Measurement Yearbook concentrates on evaluating psychological constructs while other aspects of well-being are macro in nature (e.g. ecology). However, the search in econometrics and economic journals for well-being instruments resulted in 7 entries only and again no comprehensive inventory could be found.

It is not clear why Europeans pay more attention to developing comprehensive well-being scales than their American counterparts. One plausible explanation is that the enlargement of the EU results in extensive cultural diversity and economic disparities among EU nations. Hence, scholars who could influence policies and political decision makers are motivated to obtain more precise measures in order to reach economic integration and social cohesion, and reduce differences between European countries (Somarriba & Pena, 2008). On the other hand, the *Occupy Wall Street* movement (2011), originated in the US, suggested that Americans tend to be obsessed with GDP and thus many Americans perceive that there is a magical connection between GDP and personal happiness. Although the above is not the official statement released by the movement, the popularity of this movement reflects a strong sentiment against the greed and the endless pursuit of economic growth in the US society. However, the quest of going beyond GDP has not been moved from the grass-root movement into the US policy-making arena.

2.1 GDP Is Insufficient

The value of GDP is not in question. Without sufficient wealth it is difficult for a country to meet the basic needs of citizens or build infrastructure, much less pursue higher artistic and cultural goals. World Values Survey (World Values Study Group, 1994), conducted with representative samples of approximately 1,000 participants per nation between 1990 and 1993, revealed that the correlation between life satisfaction and purchasing power parity (PPP), in which the purchasing power is adjusted by currency exchange rates, was .62. The finding that wealthier countries have higher levels of reported well-being has been confirmed in other studies (Diener & Suh, 1999). It is not surprising that an economically advantaged country would be better able to fulfill basic human needs for food, shelter, and health, as well as to have a stronger human-rights record (Diener, 2000). However,

Diener (2000) warned that in some societies higher productivity could decrease well-being if it requires long hours of boring work, high levels of stress, and little leisure time. As a matter of fact, Americans have less leisure time than Germans, Italians, British, French, and Finns due to the work hours of Americans being substantially longer than their European counterparts (OECD, 2009).

GDP mainly measures production output expressed in terms of monetary values. Unfortunately, it has often been misused as if it were a measure of well-being. This misleading indication might eventually lead to misguided policy decisions (Stiglitz, Sen, & Fitoussi, 2009). In a similar vein, Cunningham (2010) pointed out that GDP should be treated as the means or the enabler of multiple end points, rather than the sole end in itself. During the Reagan, Bush, and Clinton administrations (1980-2000), GDP per capita in America had increased 55%, and thus most Americans perceived that life was better off. However, a different story emerges when Redefining Progress developed Genuine Progress Indicator (GPI) as an alternate measure. GPI takes personal consumption, the value of household work, net fixed investment, the value of services of consumer durables, the cost of commuting, the loss of wetlands, the depletion of non-renewable resources, and several other social costs into account. It was found that during the same period of time the GPI per capita had no improvement at all (Deutsche Bank Research, 2006).

2.2 European Lead

Multiple suggestions have been made for indications of national well-being beyond GDP by different parties. Europeans are taking the lead in the movement of going beyond GDP. For example, in 2007 France's President Sarkozy, aware of the limitations of GDP, set up a Commission on the Measurement of Economic Performance and Social Progress in order to look for alternative ways of setting and measuring the goals of France. The Commission held its first meeting in 2008 and a report was compiled a year later (Stiglitz, Sen, & Fitoussi, 2009). The Commission suggested taking the following key dimensions into account in an attempt to obtain a comprehensive overview of national well-being: material living standards (income, consumption and wealth), health, education, personal activities including work, political voice and governance, social connections and relationships, environment (present and future conditions), and insecurity of an economic as well as a physical nature. In addition to objective indicators of well-being, the Commission also suggested including subjective measures of the quality-of-life, which encompasses cognitive evaluations of one's life, happiness, satisfaction, positive emotions (e.g. joy and pride), and negative emotions (e. g. pain and worry). However, the commission's general position was to avoid formulating definitive turnkey proposals on various issues regarding measuring well-being; rather, the objective of the report was to stimulate further debate.

In 2009 the European Commission published a roadmap for developing new measurement strategies to reflect the real prosperity and wellbeing of nations beyond GDP, such as including ecological and carbon footprints. According to the EU report, these initiatives are in alignment with public opinion as European citizens expecting balanced progress. Specifically, more than two thirds of EU citizens who responded to a 2008 Eurobarometer survey declared that social and environmental indicators should be on par to economic measures while evaluating progress. Only about one-sixth of the respondents prefer measurement based mainly on economic indicators to other well-being measures (Commission of the European Communities, 2009). Interestingly, the idea of including ecological and carbon footprints is in line with the Happy Planet Index (HPI) developed by the Britain-based New Economics Foundation.

Indeed, before the French Commission and the EU Commission reports, the debate of certain constructs of well-being has been ongoing for many years. Like Cunningham and President Sarkozy, several American economists (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004) also agreed that the goal of public policy should not be to maximize GDP alone; rather, a better measure of well-being should be in place in order to inform policy. However, they realized that measures of happiness are highly influenced by the immediate context along with many other factors, such as comparisons with other people and with past experiences. Cultural differences also play a crucial role in measuring well-being. In a Euro-barometer survey, 64 percent of the Danes reported that they were "very satisfied" with their lives, but only 16 percent of the French said so. This large difference might be due to different cultural orientations in the perception of life events, thus challenging the validity of the survey (Kahneman et al., 2004). As a remedy, Kahneman et al. (2004) proposed a new construct, namely, "objective happiness," and a new method for measuring it. However, Alexandrova (2005) argued that "objective happiness" cannot be a general measure of happiness, which should be subjective in nature. Instead, they insisted that the construct of happiness should be measured by Subjective Well-Being Instrument (SWBI), which is a scale for evaluating personal satisfaction of life. However, SWBI should not be considered a global measure. Keyes, Shmotkin, and Ryff (2000) found that SWBI and Psychological Well-Being Instrument (PWBI), which entail the perception of engagement with existential challenges of life, measure

related but distinct factors of well-being. In short, there is no consensus on a universal measure of well-being in terms of happiness and satisfaction that can be well-applied across all cultures and nations.

While many researchers and international organizations agree that sustainable development, a pattern of using resources to meet human needs while preserving the environment for future generations, is an important aspect of national well-being, at present different countries use different indicators. In order to extract the common ground of 28 existing indicators of sustainable development, in 2007 the United Nations, OECD, and the Statistical Office of the European Communities (Eurostat) formed the Working Group on Statistics for Sustainable Development. The mission of the group is to analyze indicator sets from twenty European countries, two non-European countries (Australia and Canada), and two international institutions (the European Union and the United Nations). It is noteworthy that the input from the United States was absent from the analysis. A year later the Working Group ((UNECE/OECD/Eurostat, 2008) released a report suggesting a smaller set of sustainable development indicators, such as temperature deviations from normal, ground-level ozone and fine particle concentrations, quality-adjusted water availability, fragmentation of natural habitats, reserves of energy resources, reserves of mineral resources, timber resource stocks, and marine resource stocks. However, the Working Group explicitly stated that this small set of indicators is offered in an exploratory fashion only. It is not intended as an international recommendation.

It might take years, if not decades, for international organizations to establish a set of universal measures of well-being, let alone psychometrically validate them. Nevertheless, exploratory research can still be carried out with incomplete and imperfect information. As mentioned before, Happy Planet Index has implemented certain European ideas, such as taking ecological factors into consideration. In order to examine whether higher international test scores, which are said to reflect more powerful cognitive capabilities, could bring about a higher degree of national well-being, this project utilized available international data found in Human Development Index (HDI) and Happy Planet Index (HPI).

3. Method

3.1 Data Sources

As an alternative to GDP, this study proposes two other indexes of national well-being, Human Development Index (HDI) and Happy Planet Index (HPI), for a richer picture of quality of life. HDI was first introduced in 1990 by the United Nations Development Project (2010). HDI goes beyond GDP to a broader definition of well-being by providing a composite measure of three dimensions of human development: living a long and healthy life (measured by life expectancy), education (measured by years of schooling), and a decent standard of living (measured by GNI per capita). The 2010 report, which is composed of statistics collected from 169 nations, was downloaded and used in this study. GNI per capita is also a measure of national wealth, and as previously mentioned, the relationship between international test scores and national wealth has been well-studied. Based upon life expectancy and years of schooling, UN also developed the non-income HDI index. At first glance, it may be difficult to conceptualize the link between test scores and schooling as they seem to be two concurrent aspects of education rather than causally related. Nonetheless, when students have better academic performance, as measured by PISA test scores at age 15, it can be inferred that there will be more qualified candidates for continuous education beyond Grade 10 or 11, resulting in a society consisting of a well-informed population. One may argue that although better math and science education could lead to improved medical technology and healthcare, eventually extending life expectancy, the benefits could only be seen in the indefinite future. Nevertheless, advancements in medical technology and public health could at least prevent some premature deaths, and therefore it is still meaningful to include life expectancy as the outcome measure.

Happy Planet Index, developed by the New Economics Foundation (2009), is an efficiency ratio to indicate how well a nation manages its resources to achieve well-being in terms of long, happy, and meaningful lives. HPI is a composite score of three measures: life expectancy, life satisfaction, and ecological or environmental footprint (EF). Life satisfaction is measured by asking the following type of question: "All things considered, how satisfied are you with your life as a whole these days?" By combining life satisfaction and life expectancy, another index called happy life years (HLY) was introduced in order to ensure that both the subjective and objective elements of well-being are captured. Specifically, a happy life is not desirable if it is very short, but neither is a long life if it is miserable. The ecological footprint is a measure of human demand on natural resources by comparing human demand with ecological capacity to regenerate. The higher the EF is, the more resources an individual or society uses. HPI was launched in 2006 and the 2009 report was compiled with the data collected in 143 nations.

PISA (OECD, 2010a) administers tests to 15-year-old students recruited from 30 countries every three years (2000, 2003, 2006, and 2009). PISA is composed of mathematics, science, and reading tests. This study included only math and science test scores due to the controversy over alleged cultural bias in the PISA reading test (Bracey, 2005). Because it was found that PISA science and math scores collected in 2000 have a very strong correlation ($r = .945$), a composite score was used in this study.

It is important to note that while it is a common practice for studies on PISA and GDP to use concurrent data (e.g. Chen & Luoh, 2010; Cheung & Chan, 2008), this approach is questionable. OECD (2010b) realized that concurrent analysis is based on the assumption that the average scores observed for students are a good proxy of labor force skills. This assumption is tied to another assumption that the educational outcomes within the same countries remain roughly constant. However, this assumption is false, and thus results in measurement errors. Indeed, GDP and other measures of well-being are usually unresponsive to short term changes of investment in education and research; a concurrent effect is not to be expected. Given the time it would take for a student who took the PISA tests to complete their studies and (potentially) contribute to national development, this study utilized the earliest PISA data set (2000) along with the latest HDI (2010) and HPI (2009). One may argue that by 2009 or 2010, those 15-year old students who took PISA tests were 24 or 25-years old only and thus their contribution to the society might not be significant enough to improve the national well-being. Using TIMSS data would make it worse because TIMSS 1999 tested students at Grade 8 (12-13 years old) and TIMSS 1995 was the first test, which was not thoroughly validated. In other words, despite its shortcoming, PISA 2000, HDI 2010, and HPI 2009 are by far the best data available. The author was aware of this limitation and thus the results were interpreted with caution.

3.2 Analysis

This study employed exploratory factor analysis and principle component analysis to examine the factor structure of HDI and HPI. For exploratory purposes, Pearson's *r*, scatterplot matrix, and simple regression were used to investigate the inter-relationships among the independent and dependent variables. Due to the fact that pairwise correlations and simple regression models yield very similar numeric outputs, the focus of regression was placed on data visualization instead of repeating redundant information.

A simple regression modeling without taking moderators and other grouping factors into account might lead to biased results. However, there could be many potential moderators and grouping factors, such as the technological/economic level of development between countries and cultural backgrounds. The model will be overly complicated if three or more such variables are included into the model. In order to obtain a single moderating variable crystalized from the cultural, political, and geographical profile of the nations, two-step cluster analysis was utilized to classify these nations in terms of HDI and HPI. Finally, regression analysis conditioning on the moderator was performed.

Cluster analysis is essentially a data reduction method. Conceptually and procedurally speaking, cluster analysis can be viewed as Analysis of Variance (ANOVA) or Analysis of Covariance (ANCOVA) in reverse. In ANOVA and ANCOVA the data are analyzed according to the pre-determined grouping factors and covariates. When many independent variables must be taken into consideration, the model might be very complicated. For example, if it is necessary to include five factors and each factor has three levels, the proposed model would be a 3 X 3 X 3 X 3 X 3 ANOVA. Needless to say, it will be difficult to interpret the results yielded from such a model. Conversely, cluster analysis condenses the grouping factors and covariates by proposing some initial clusters, and then moving observations between those clusters until the variability within the same cluster is minimized and the between-group variability is maximized. It is equivalent to ANOVA in reverse because the significance of ANOVA is based on the *F* value, which is the ratio of between-group variability and within-group variability (Hill & Lewicki, 2005).

In addition, cluster analysis is considered a data reduction method in the sense that it is complementary to factor analysis. Factor analysis groups variables based on the response patterns of the observations whereas cluster analysis groups observations or cases based on the variables of interest (Antonenko, Toy, & Niederhauser, 2012). However, unlike factor analysis that follows certain rules (e.g. eigenvalue => 1), cluster analysis is example-based (Witten, Frank, & Hall, 2011). As mentioned before, cluster analysis starts with initial clusters (examples) as the reference and thus this is truly data-driven.

There are several clustering algorithms, namely, *K*-means, Hierarchical, and Two-step. In this study Two-step clustering was chosen for its efficiency and the good match between the data structure and the method. As the name implies, Two-step clustering consists of two steps. Step one is called pre-clustering, in which a cluster feature tree is constructed by scanning all observations. When a case is scanned, the algorithm utilizes the log

likelihood distance measure to determine whether the observation should be merged with other observations or form a new cluster. In the second step the hierarchical clustering method is applied to the pre-clusters and then suggests a set of possible solutions (Yu, 2010). At the end the best solution is chosen out of the set according to the Akaike Information Criterion (AIC) (Akaike, 1973) or the Bayesian Information Criterion (BIC) (Schwarz, 1978). In alignment to the principle of data reduction, both criteria aim to yield a parsimonious solution by imposing a heavy penalty on complicated models. In addition, K-means and Hierarchical clustering methods allow for continuous-scaled data only. On the contrary, two-step clustering is a more versatile procedure because it can accept both categorical and continuous-scaled data and both data types are included in this study.

4. Results

4.1 Factor Structure

Exploratory factor analysis based upon the method of maximum likelihood suggested that there is no unique solution for the factor structure of HDI and HPI, and the null hypothesis that one factor is sufficient is rejected in both cases. This conclusion was verified by principal component analysis. Table 1 and 2 show that not all variables could be loaded into a single principal component. It is beyond the scope of this paper to discuss the psychometric properties of HDI and HPI. Simply put, although the author followed the developers of HDI and HPI to treat these scales as composite indexes due to mathematical convenience and conceptual simplicity, it is noteworthy that indeed HDI and HPI carry multiple dimensions. Hence, both the composite scores of HDI and HPI, as well as some of their sub-categories, were examined later.

Table 1. Loadings and eigenvectors of HDI

Eigenvectors	Component 1	Component 2	Component 3
Life expectancy 2010	0.589427	-.498553	0.635626
Mean years of schooling 2010	0.522223	0.835481	0.171041
GNI per capita 2010	0.616327	-.231122	-.752811

Table 2. Loadings and eigenvectors of HPI

Eigenvectors	Component 1	Component 2	Component 3
Life satisfaction 2009	0.602105	-.001378	-.798416
Life Expectancy 2009	0.564473	0.707958	0.424461
Environmental footprint 2009	0.564659	-.706254	0.427043

4.2 Descriptive Statistics

Table 3 shows the descriptive statistics of all the variables. HPI has life expectancy data of 2009, but this was superseded by the data of 2010 available in HDI, and thus the former was excluded from the table.

Table 3. Descriptive statistics of all variables

Variable	Mean	Std Dev	Minimum	Maximum
Environmental footprint 2009	5.53374	1.84271	2.35623	10.19218
GNI per capita 2010	30781	11325	10607	58810
Happy life years 2009	56.76062	7.69052	38.13156	64.57039
HDI 2010	0.85046	0.05767	0.69856	0.93756
HPI 2009	41.90677	6.78005	28.47192	61.01450
Life expectancy 2010	79.06983	3.45766	67.20800	83.16600
Life satisfaction 2009	7.26292	0.78870	5.43127	8.14400
Non-income HDI value	0.88013	0.06136	0.72828	0.98907
PISA composite	985.63333	84.48198	709.00000	1107
Years of schooling 2010	10.63260	1.36026	7.17790	12.63130

Table 4 shows the correlation matrix of the composite measures. It suggests that PISA composite index is strongly correlated with HDI and non-income HDI, but not HPI or HLY. Table 5 and 6 depict the correlation matrix of PSI and the individual variables of both scales, respectively.

Table 4. Correlation matrix of PISA and composite variables in HDI and HPI

	HDI 2010	Non-income HDI 2010	HPI 2009	HLY 2009
PISA composite 2000	0.74367	0.74020	-0.28171	0.31166
	<.0001*	<.0001*	0.1315	0.0936
HDI 2010		0.95845	-0.29384	0.69126
		<.0001*	0.1150	<.0001*
Non-income HDI 2010			-0.26318	0.61523
			0.1600	0.0003*
HPI 2009				0.13577
				0.4744

* Significance at .05 level

Table 5. Correlation matrix of PISA and individual variables of HDI

	Life expectancy 2010	Years of schooling 2010	GNI per capita 2010
PISA composite 2000	0.53688	0.66122	0.46315
	0.0022*	<.0001*	0.0100*
Life expectancy 2010		0.41350	0.66281
		0.0231*	<.0001*
Years of schooling 2010			0.50221
			0.0047*

* Significance at .05 level

Table 6. Correlation matrix of PISA and individual variables of HPI

	Life satisfaction 2009	Environmental footprint 2009	Life expectancy 2010
PISA composite 2000	0.16854	0.29942	0.52523
	0.3733	0.1080	0.0029*
Life satisfaction 2009		0.57395	0.57347
		0.0009*	0.0009*
Environmental footprint 2009			0.46402
			0.0098*

* Significance at .05 level

4.3 Preliminary Regression Modeling

Although the preceding scatterplot matrices were constructed to investigate the inter-relationships among the variables, only some of the pairwise scatterplots that are related to the research focus were shown in the subsequent analysis. Figures 1 and 2 depict the relationships between PISA and HDI, and between PISA and non-income HDI, respectively.
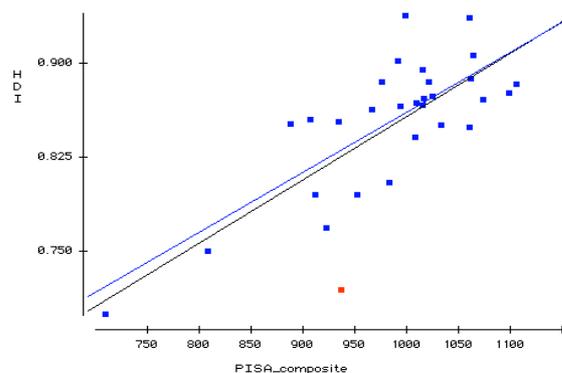

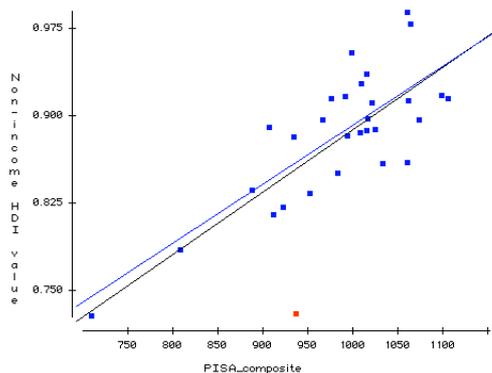Figure 1. Scatterplot of PISA in 2000 and HDI in 2010

Figure 2. Scatterplot of PISA in 2000 and non-income HDI in 2010

In both graphs, the red dot at the bottom of the Y-axis denotes Russia. The lower black regression line includes all observations whereas the upper blue regression line is plotted without Russia. At first glance Russia seems to be an outlier. Nevertheless, in the regression model of PISA against HDI, the changes of the parameter estimate, the variance explained (from R^2=.55 to R^2=.60), and the regression slope are minimal even if Russia is excluded from the model. As shown in Table 7 and Table 8, both models lead to the same conclusion that PISA composite is a significant predictor of HDI.

Table 7. Regression models of HDI against PISA with Russia

Parameter Estimates with Russia					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.35007	0.08531	4.10	0.0003
PISA composite	1	0.00050768	0.00008625	5.89	<.0001

Table 8. Regression models of HDI against PISA without Russia

Parameter Estimates without Russia					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.37937	0.07421	5.11	<.0001
PISA composite	1	0.00048176	0.00007490	6.43	<.0001

Similarly, while regressing PISA against non-income HDI, no matter whether Russia is present or absent from the data set, there is no substantial change in the model. The variance explained increases from .54 to .61, and the predictive power of PISA for non-income HDI remains strong.

Table 9. Regression models of non-income HDI against PISA with and without Russia

Parameter Estimates with Russia					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.35026	0.09128	3.84	0.0006
PISA composite	1	0.00053759	0.00009229	5.83	<.0001
Parameter Estimates without Russia					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.38447	0.07638	5.03	<.0001
PISA composite	1	0.00050733	0.00007709	6.58	<.0001

Figure 3 and 4 illustrate the relationship between PISA and HPI, and between PISA and HLY, respectively. The two red dots located in the upper left corner of the scatterplot represent Brazil and Mexico. Obviously, these two outliers substantially affect the slope, resulting in a misleading conclusion. In the model of regressing HPI against PISA, the slope changes from negative to positive when the two extreme cases are removed. By the same token, when the outcome variable is HLY, there is a drastic shift in the slope without Mexico and Brazil. Table 10 illustrates the regression models with and without the two outliers. The finding suggests that PISA in 2000 is a significant predictor of HPI and HYI in 2009.

Figure 3. Scatterplot of PISA in 2000 and HPI in 2009   Figure 4. Scatterplot of PISA in 2000 and HLY in 2009

Table 10. Regression models of HPI against PISA with and without Mexico and Brazil

Parameter Estimates with Mexico and Brazil					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	64.19007	14.39419	4.46	0.0001
PISA composite	1	-0.02261	0.01455	-1.55	0.1315
Parameter Estimates without Mexico and Brazil					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.02914	15.97373	0.13	0.8999
PISA composite	1	0.03864	0.01592	2.43	0.0225

Table 11. Regression models of HLY against PISA with and without Mexico and Brazil

Parameter Estimates with Mexico and Brazil					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	28.79716	16.16875	1.78	0.0858
PISA composite	1	0.02837	0.01635	1.74	0.0936
Parameter Estimates without Mexico and Brazil					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.64827	24.13979	-0.11	0.9135
PISA composite	1	0.05933	0.02406	2.47	0.0206

The same patterns could be found in the regression models of PISA as the predictor and the sub-categories as the outcome variables (e.g. life satisfaction, environmental footprint, life expectancy years of schooling, and GNI per capita). For clarityof illustration, only one scatterplot and one set of regression lines were shown in Figure 5. Based on this information, Mexico and Brazil were excluded from subsequent analysis.
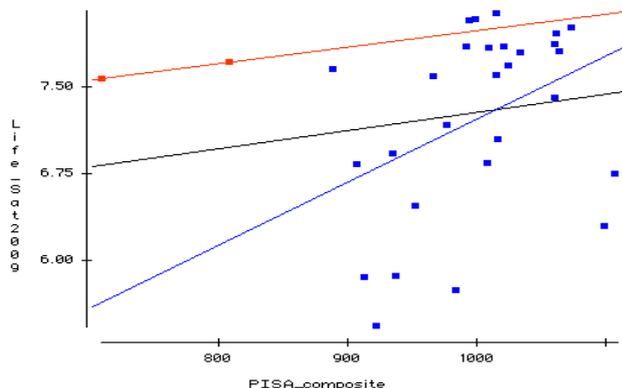

Figure 5. Scatterplot of PISA in 2000 and life satisfaction in 2009

4.4 Two-step Cluster Analysis

Competency in math and science, as measured by PISA, is a significant predictor of national well-being, as measured by HDI and HPI, when outliers (Brazil and Mexico) are excluded from modeling. However, this analysis is insufficient if technological, cultural, and geopolitical factors are not taken into account. Thus, a two-step cluster analysis was conducted in an attempt to group these nations by their common threads. When composite indexes of well-being (HDI, HPI, HLY, and non-income HDI) were used in the classification, it was found that the most important predictors for group classification are HLY and HDI, as shown in Figure 6. When sub-categories were used as the criteria of classification, life satisfaction and life expectancy were considered the most important (see Figure 7).
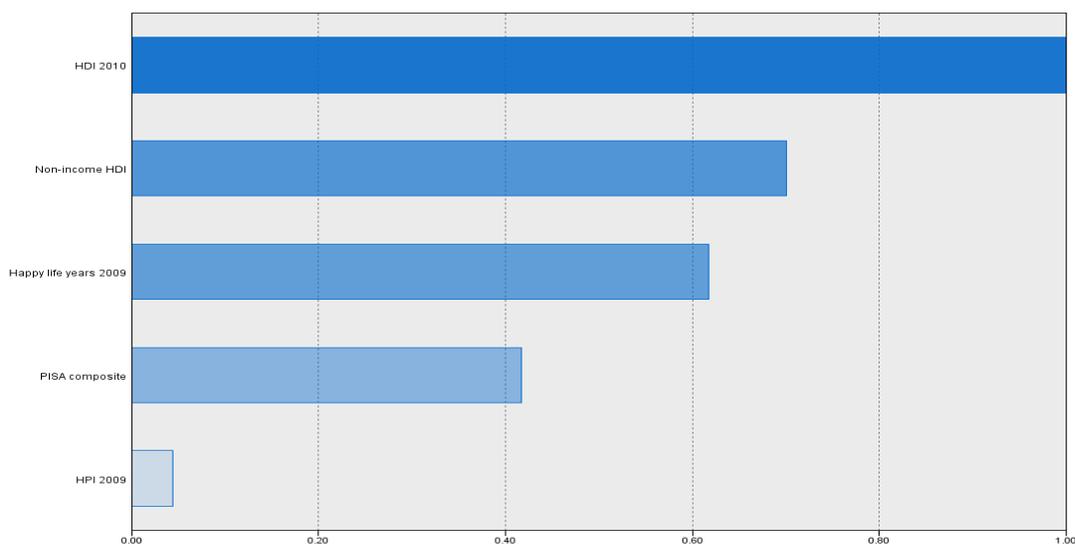

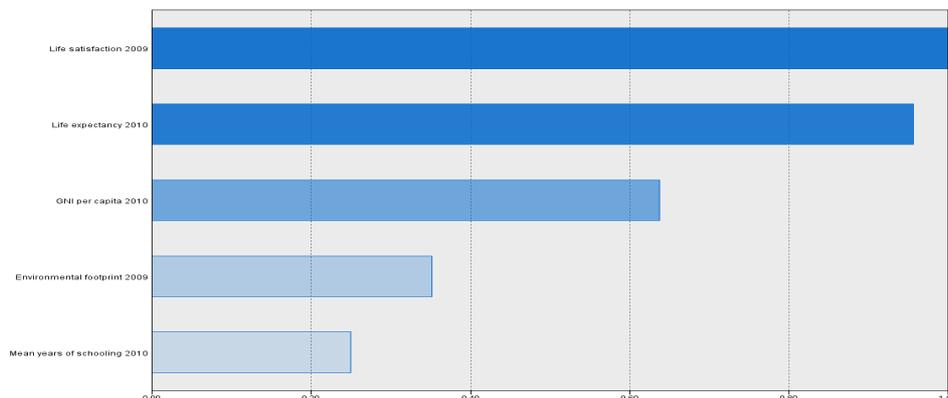Figure 6. Ranked importance of composite indexes as predictors in cluster analysis

Figure 7. Ranked importance of sub-categories as predictors in cluster analysis

The clustering procedures using composite scores and sub-categories yielded the same result. Table 11 shows the two sub-groups and their members. The mean scores of all well-being measures of Cluster 1 are higher than those of Cluster 2. However, it is important to point out that a higher ecological footprint score denotes more resource use. Thus, Cluster 1 is doing better than Cluster 2 in almost all aspects of well-being except ecological footprint. Four of the five nations (Hungry, Latvia, Poland, Portugal, and Russia) in Cluster 2 are post-Communist countries, three of them are located in East Europe (Russia spans across East Europe and Asia), and one of them is situated in South Europe. In Cluster 1, only one (Czech Republic) is an East European nation.

Table 11. Subgroups from cluster analysis

Cluster	1	2
Size in percentage	82.1%	17.9%
Size in count	23	5
Mean of HLY	59.87	42.65
Mean of HDI	0.88	0.78
Mean of non-income HDI	0.91	0.81
Mean of HPI	41.32	38.04
Mean of life satisfaction	7.53	5.87
Mean of life expectancy	80.58	73.84
Mean of GNI per capita	35,360.00	17,116.54
Mean of ecological footprint	6.13	3.84
Mean of years of schooling	11.05	8.80
Countries	Australia	Hungary
	Austria	Latvia
	Belgium	Poland
	Canada	Portugal
	Czech Republic	Russia
	Denmark	
	Finland	
	France	
	Germany	
	Greece	
	Iceland	
	Ireland	
	Italy	
	Japan	
	Korea	
	Luxembourg	
	New Zealand	
	Norway	
	Spain	
	Sweden	
	Switzerland	
	United Kingdom	
	United States	

4.5 Regression by Cluster

When cultural and geopolitical factors were taken into accunt by using cluster as a moderating variable, some interesting patterns were discovered. When all observations were treated as one group, the regression shope of PISA against HLY was steep (see Figure 8). However, when separate regression lines were fitted into the two subsets, the slope of the better-off nations was almost flat, meaning that increase in the PISA test scores does not tend to substantially increase happiness when a nation has already reached a high level of development. For Cluster 2 nations, the slope was not as steep as the overall regression model. In other words, the strong effect vanished after clustering. The same pattern was also found in the regression models when using life satisfaction, life expectancy, and HDI as the outcome variables (see Figure 9-12). Figure 9 shows that in Cluster 1 there was virtually no relationship between life satisfaction and PISA while in Cluster 2 the slope was positive. Figure 10 displays that Cluster 2 regression line resembled the overall regression line, but the Cluster 1 line did not. Figure 11 depicts the relationships between PISA and life expectancy. Again, the regression lines of Cluster 1 and Cluster 2 went to opposite directions. In the model of regressing PISA against GNI per capita, the overall regression line was positive, but the regression slopes of both Cluster 1 and 2 were negative (see Figure 12). It is important to point out that not only does the relationship between GNI per capita and the better-off nations became negative, but also this pattern was present among Cluster 2 nations. This is a typical example of Simpson's Paradox, in which the conclusion yielded from the aggregate data is opposite to that drawn from the partitioned data.
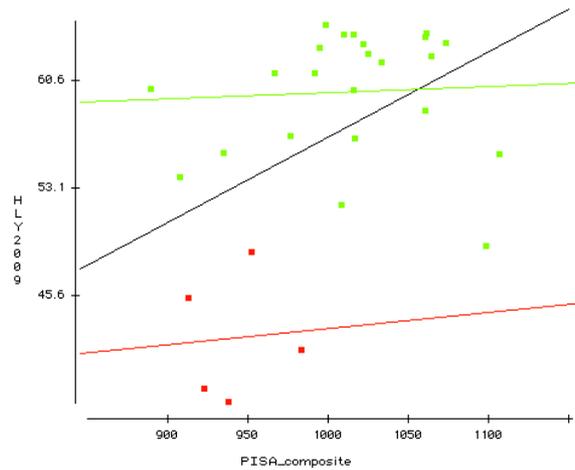

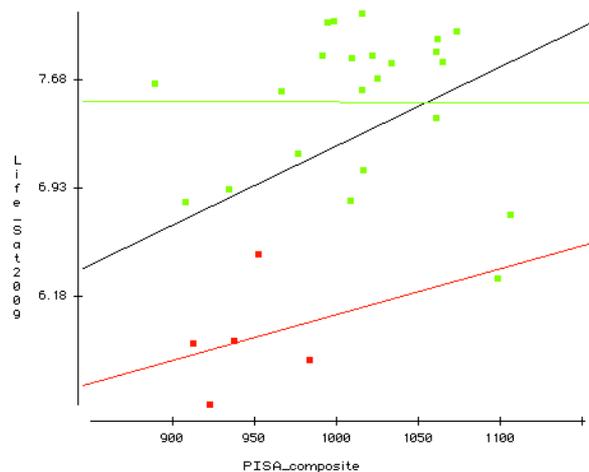Figure 8. Regression model of PISA against HLY conditioning on cluster

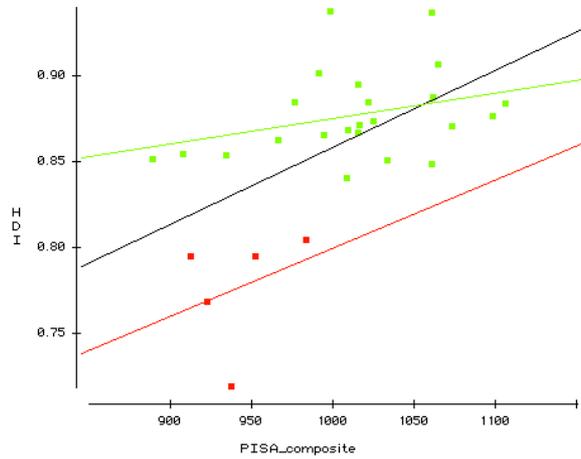Figure 9. Regression model of PISA against life satisfaction conditioning on cluster

Figure 10. Regression model of PISA against HDI conditioning on cluster
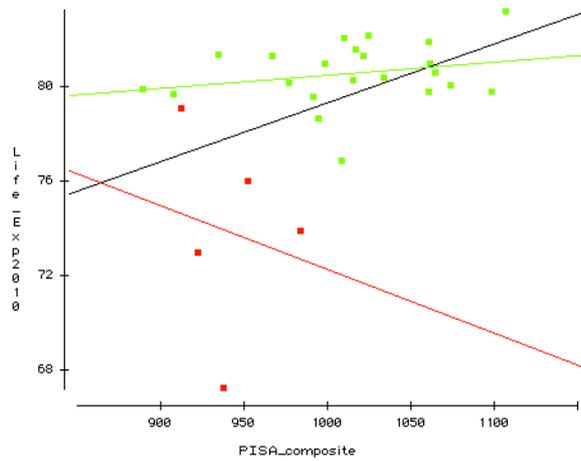

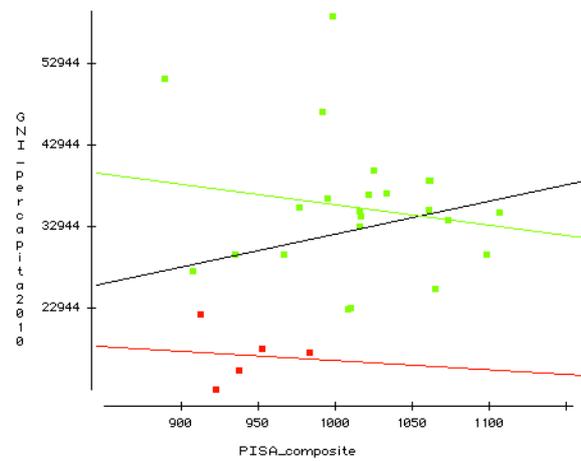Figure 11. Regression model of PISA against life expectancy conditioning on cluster

Figure 12. Regression model of PISA against GNI per capita conditioning on cluster

5. Discussion

5.1 Summary of Findings

This study attempts to go beyond GDP to examine whether higher international test scores, which are said to indicate higher national intelligence, might lead to better national well-being as measured by HDI and HPI. Initial analyses show that the quality of life measured in 2009 and 2010 is significantly associated with PISA science and math test scores in 2000. However, cluster analysis suggests that there is a clustering pattern among these PISA participating nations in terms of HDI and HPI. Four out of five nations that have lower measures of national well-being (except ecological footprint) are post-Communist nation whereas only one out of 23 nations that have a higher quality of life belongs to this category. This implies that there might be a cultural and geopolitical dimension lurking in the background. When clusters are involved in the conditioning regression models, the relationships between PISA, HDI, and HPI become unstable and inconsistent. Thus, the author withholds the generalization that PISA scores collected about a decade prior could predict those national well-being scores, and that people who possess higher cognitive skills reflected by higher test scores are able to make smart choices regarding time allocation for a balanced and happy life, resource use for sustainable development, and other aspects of well-being. Further research on this topic must be localized by regions or cultural classifications.

5.2 Limitations

Nonetheless, this study has certain limitations. First, the data sets used in this study are national-level summaries, not individual records. No doubt the conclusion is subject to the ecological fallacy, which consists in thinking that relationships found in groups necessarily hold for individuals (Freedman, 1999), and thus readers should interpret the results with caution. Second, national well-being is an open concept that is hardly operationalized by HDI or HPI. As mentioned before, international organizations have not yet established a common set of indicators of well-being, and different researchers hold different views toward happiness as a psychological construct. As mentioned repeatedly, these initiatives are mostly European endeavors based on Euro-barometer polls. At the time of this writing, certain international task forces, such as the Working Group on Statistics for Sustainable Development, are operating without the input from the US. Therefore the goal of formulating a truly cross-cultural measurement scale of well-being awaits more cross-cultural dialogs and collaboration.

5.3 Future Directions and Recommendations

The overarching principle of this analysis is the rejection of overdependence on one single outcome, namely, economic output. This analysis is based on the European initiatives of going beyond GDP in measuring progress and national well-being. However, even if a researcher would wish to confine the research goal to examining the relationship between international test scores and economic performance, GDP is still too narrow to reflect the economic well-being of a nation. As a remedy, World Economic Forum (Schwab, 2010) has been using twelve pillars to measure economic competitiveness, namely, institutions (legal and administrative framework), infrastructure, macroeconomic environment, health and primary education, higher education and training, goods market efficiency, labor market efficiency, financial market development, technological readiness, market size, business sophistication, and innovation. Similarly, the European Union had developed the European Innovation Scoreboard (EIS), which includes 19 indicator statistics for tracking investment and 10 measures of output or impact (Pro INNO Europe, 2010). These statistics have been expanded to develop the Global Innovation Scoreboard (GIS) (Cunningham, 2010).

This exploratory study examines just a few selected aspects of national well-being and thus the findings are far from conclusive. The author recommends that in the future various well-being measures should be rigorously validated by psychometric methods, and also that the relationship between PISA scores and other measures of national well-being should be broadly and deeply examined by religion or even by country. This might result that educational policy-making could be better informed by a cluster of indictors, rather than focusing on GDP, PPP, or GNI per capita alone.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *International Symposium on Information Theory* (pp. 267-281). Budapest: Akademia Kiado.

Alexandrova, A. (2005). Subjective well-being and Kahneman's "objective happiness". *Journal of Happiness Studies, 6*, 301-324. http://dx.doi.org/10.1007/s10902-005-7694-x

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60,* 383-398. http://dx.doi.org/10.1007/s11423-012-9235-8

Bicchieri, C. (1993). *Rationality and coordination*. Cambridge, England: Cambridge University Press.

Bracey, G. (2005). Put out over PISA. *Phi Delta Kappan, 86*(10), 797-798.

Buros Institute of Mental Measurements. (2012). *Mental measurement yearbook*. Lincoln, NE: Author.

Chen, S. S., & Luoh, M. C. (2010). Are mathematics and science test scores good indicators of labor-force quality? *Social Indicators Research, 96*, 133-143. http://dx.doi.org/10.1007/s11205-009-9470-5

Cheung, H. Y., & Chan, A. (2008). Understanding the relationships among PISA scores, economic growth and employment in different sectors: A cross-country study. *Research in Education, 80*(1), 93-106.

Commission of the European Communities. (2009). GDP and beyond: Measuring progress in a changing world. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=com:2009:0433:FIN:EN:PDF

Cunningham, P. (2010). European Union: Measuring success. *Issues in Science and Technology*. Retrieved from http://www.issues.org/26.3/cunningham.html

Deutsche Bank Research. (2006). Measures of well-being: There is more to it than GDP. Retrieved from http://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD0000000000202587.pdf

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American psychologist, 55*, 34-43. http://dx.doi.org/10.1037//0003-066X.55.1.34

Diener, E., & Sub, E. (1999). Societies we live in: International comparisons. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 434-452). New York: Russell Sage Foundation.

Freedman, D. (1999). Ecological inference and the ecological fallacy. Retrieved from http://www.stanford.edu/class/ed260/freedman549.pdf

Hanushek, E., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*, 607-668. http://dx.doi.org/10.1257/jel.46.3.607

Hill, T., & Lewicki, P. (2005). *Statistics: Methods and applications*. Tulsa, OK: StatSoft, Inc.

Kahneman, D., Krueger, A., Schkade, D., Schwarz, N., & Stone, A. (2004). Toward national well-being accounts. *American Economic Review, 94*, 429-434. http://dx.doi.org/10.1257/0002828041301713

Keyes, C., Shmotkin, D., & Ryff, C. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology, 82*, 1007-1022. http://dx.doi.org/10.1037//0022-3514.82.6.1007

Lynn, R., & Mikk, J. (2007). National differences in intelligence and educational attainment. *Intelligence, 35,* 115-121. http://dx.doi.org/10.1016/j.intell.2006.06.001

New Economics Foundation. (2009). *The Happy Planet Index 2.0: Why good lives don't have to cost the earth*. Retrieved from http://www.happyplanetindex.org/public-data/files/happy-planet-index-2-0.pdf

Occupy Wall Street. (2011). America's obsession, GDP, and world domination. Retrieved from http://occupywallst.org/forum/americas-obsession-gdp-and-world-domination/

OECD. (2010a). *Database: PISA 2000*. Retrieved from http://pisa2000.acer.edu.au/

OECD. (2010b). *The high cost of low educational performance: The long run economic impact of improving PISA outcomes*. Retrieved from http://www.oecd.org/dataoecd/11/28/44417824.pdf

Organization for Economic Cooperation and Development (OECD). (2009). *Growing unequal? Income distribution and poverty in OECD countries*. Paris: Author.

Pro INNO Europe. (2010). European Innovation Scoreboard (EIS) 2009. Retrieved from http://www.proinno-europe.eu/page/european-innovation-scoreboard-2009

Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence, 36*, 127-142. http://dx.doi.org/10.1016/j.intell.2007.02.002

Schwab, K. (2010). The global competitiveness report 2010-2011. In *World Economic Forum*. Retrieved from http://www.weforum.org/en/initiatives/gcp/Global%20Competitiveness%20Report/index.htm

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464. http://dx.doi.org/10.1214/aos/1176344136

Somarriba, N., & Pena, B. (2008). Synthetic indicators of quality of life in Europe. *Social Indicators Research, 94*, 115-133. http://dx.doi.org/10.1007/s11205-008-9356-y

Stiglitz, J., Sen, A., & Fitoussi, J. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. Retrieved from http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

UNECE/OECD/Eurostat. (2008). Report on measuring sustainable development: statistics for sustainable development, commonalities between current practice and theory. Retrieved from www.unece.org/stats/archive/03.03f.e.htm

United Nations Development Project. (2010). Human development report 2010. Retrieved from http://hdr.undp.org/en/statistics/

West, M. (2011). Education and global competitiveness: Lessons for the U.S. from international evidence. In *American Enterprise Institute*. Retrieved from http://www.aei.org/files/2011/09/29/West%20paper%2009-19-11.pdf

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3th ed.). Boston, MA: Morgan Kaufman.

World Values Study Group. (1994). *World Values Survey, 1981-1994 and 1990-1993*. Ann Arbor, MI: Institute for Social Research.

Yu, C. H. (2010). Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research, 3*(1), 9-22. Retrieved from http://mvint.usbmed.edu.co:8002/ojs/index.php/web/article/download/455/460