Cross-country Analysis of ICT and Education Indicators: An Exploratory Study

Ahmad R Pratama¹,²

¹ Department of Informatics, Universitas Islam Indonesia, Sleman, DIY 55584 Indonesia
² Department of Technology and Society, Stony Brook University, NY 11794 USA

E-mail: ahmad.rafie@uii.ac.id; ahmad.pratama@stonybrook.edu

Abstract. This paper explores the relationship between world ICT and education indicators by using the latest available data from World Bank and UNESCO in range of 2011–2014 with the help of different exploratory methods such as principal component analysis (PCA), factor analysis (FA), cluster analysis, and ordinary least square (OLS) regression. After dealing with all missing values, 119 countries were included in the final dataset. The findings show that most ICT and education indicators are highly associated with income of the respective country and therefore confirm the existence of digital divide in ICT utilization and participation gap in education between rich and poor countries. It also indicates that digital divide and participation gap is highly associated with each other. Finally, the findings also support the idea of reverse causality between ICT and education; higher participation rate in education increases technology utilization, which in turn helps promote better outcomes of education.

1. Introduction

It is a fact that technology, particularly information and communication technology (ICT), has been growing rapidly within the past decade. While it exists in different forms from telephone, either landline and mobile, to the Internet, they all are growing at an exponential rate and have been impacting the shapes of human lives, for better or for worse. It connects people in different parts of the world, breaking the barriers of time and distance. Nevertheless, despite some negative side effects one might argue about it, it does open numerous possibilities to improve the quality of human lives. One of them is by improving the quality of education.

In this paper, we try to explore more about the relationship between ICT and education by using worldwide indicators, summarize several indicators to see if there is something in common or the point out important differences among them, examine different level of education in term of technology utilization, and see what effect technology utilization has to education outcomes. Furthermore, we will also examine if income has something to do in the relationship between ICT and education in many countries.

2. Conceptual Background

Two decades ago, long before the era of Google, Wikipedia, or Facebook, Kuntz suggested that the Internet promotes literacy [1]. Five years later, another scholar argued that ICT is an option to close the gap of educational inequalities [2]. Ever since, the notion that ICT helps improve education has...
been widely accepted with the fact of the rise of many e-learning platforms, one of which is well known as massive, open, online course (MOOC) as well as the popularity of Educational Technology as a new research area. The more recent trend also includes mobile devices where we no longer use them only to make phone calls or to send text messages with the introduction of smartphones. Nowadays, people can play games, socialize with other people across the globe, and of course educate themselves by using their own mobile devices. Many have declared their belief in the opportunities and advantages of mobile learning [3,4].

Interestingly, it seems that the relation between ICT and Education is not a one-way causality. Some other researchers argued that technology utilization is also driven by social factors, one of which is education [5, 6]. Using data from 2008-2009, the same researchers showed that tertiary enrollment rate is one of the most important factor driving technology utilization [7]. In other words, the higher tertiary enrollment rate in a country, the more people in that country will use ICT. These findings suggest that there is actually reverse causality between ICT and education and this relationship is worth to explore more, preferably with the most recent data available. Furthermore, it is also interesting to see how different the trend and the relationship in poor and rich countries is, given that digital divide has been widely accepted as a world problem for the past decade [8].

3. Data and Methodology

3.1. Variables of Interest
There are four indicators of ICT and five indicators of education with one additional indicator serving as a control variable used in the analyses. The indicators for ICT are; 1) Internet Users (per 100 people), 2) Fixed/Wired Broadband Subscriptions (per 100 people), 3) Mobile Cellular Subscriptions (per 100 people), and 4) Secure Internet Servers (per 1 million people). These indicators are selected since they represent ICT penetration and are commonly used by many world organizations such as World Bank, OECD, and ITU as key ICT indicators. Furthermore, one particular indicator (i.e. Secure Internet Servers) also represents ICT infrastructure strength of the country as it shows producer point of views rather than consumer’s as the other three do.

Meanwhile, the education indicators are divided in two main categories; participation in education, represented by 1) Primary Enrollment Rate (%), 2) Secondary Enrollment Rate (%), 3) Tertiary Enrollment Rate (%), and the outcomes of education, represented by 4) Literacy Rate (%), and 5) Scientific and Technical Journal Articles Published. For the participation in education, it is quite obvious to include all three levels of education while the two different indicators for outcomes are used to accommodate this difference in education level. The first indicator is more associated with primary and secondary education while the last one is used to measure the indicator of tertiary education. Furthermore, these two indicators are also selected as they both can be seen as direct outcomes of education, thus minimizing the potential effect of external factors that are not included in this study.

Lastly, GDP (per capita) is used as a measure of income serving as a control variable in this study. All data were taken mainly from the World Bank using the average value of any available data within 2011-2014. Some missing values for Literacy Rate were filled by using additional data from UNESCO for the same time range. After dealing with all missing values, 119 countries were included in the final dataset.

3.2. Analysis Methods
Principal component analysis (PCA) was used to answer the first research question about the pattern and directions of these ICT and education indicators. The same PCA in addition to factor analysis (FA) were then used to summarize the data. In both PCA and FA, correlation matrix was used as opposed to covariance matrix due to metric differences between variables. Furthermore, to compare the differences between poor and rich countries, PCA was repeated for two subsets, each representing poor countries (i.e. GDP per capita less than $1,047) and rich countries (i.e. GDP per capita over $12,000) in addition to cluster analysis (CA) to see whether or not poor and rich countries fall under different clusters. Lastly, ordinary least squares (OLS) regression is used to find the impact of
educational level (primary, secondary, and tertiary) on technology utilization as well as the impact of ICT indicators on educational output (journal articles and literacy rate).

As 6 out of 10 variables were either highly skewed (i.e. GDP per capita, secure internet servers, and journal articles), moderately skewed (i.e. broadband and tertiary enrollment), or negatively skewed (i.e. literacy rate), log transformation was done to deal with highly skewed distributions and square root transformation was performed on the moderately skewed variables. Meanwhile, dealing with negatively skewed distribution involved three steps transformation where a reflection was done at the first step (i.e. by subtracting it from the maximum possible value) before log transformation was applied. The final step was to reflect back the log transformed variable so it resembles the original values (i.e. by subtracting it from the maximum value of the log transformed variable).

Unfortunately, even after the transformation, the dataset still failed to pass multivariate normality test. Therefore, the analyses were then conducted with both the original as well as the transformed data and so that the results between two of them can be compared. Apparently, there was no significant difference when the analyses were done with either the original or the transformed data. However, the transformed data provide much better results in PCA, FA, and CA. The original dataset was used in OLS Regression and log transformation will be done separately in each regression model as necessary.

In term of outliers, there are three multivariate outliers detected for the transformed data (i.e. Cuba, Sudan, and Turkmenistan). However, these outliers do not affect the findings since we got very similar results whether or not the outliers are excluded from all analyses. Meanwhile for the rich countries and poor countries subsets, the transformed data did pass multivariate normality test. No outlier was detected for the rich countries and one non-influential outlier (i.e. Chad) was detected for the poor countries.

4. Findings and Discussions

The PCA loadings show that the first principal component of this dataset is associated with nine out of ten variables with evenly distributed proportion, meanings those nine variables are associated to each other very well. Furthermore, the biplots as can be seen in Figure 1 show that all four ICT indicators and four out of five education indicators go to the same direction with income. Only primary enrollment ratio goes to quite the opposite direction. This suggests that the level of technology utilization as well as the participation and the outcome of education are heavily affected by income. Richer countries tend to have more technology utilization and more participation in education as well as better outcome.

When it comes to ICT indicators, it turned out that all indicators go to relatively the same main direction. However, it is important to note that mobile subscription number has slightly different deviation compared to the other three indicators due to the fact that not all people use their mobile devices to access Internet. While it is true that Internet access from mobile devices such as smartphones and tablets has become more and more popular lately, some people are still using regular cell phones that are mainly used merely to make phone calls and to send text messages.

As for education indicators, the biplot emphasizes the previous results that primary enrollment has quite large deviation compared to all other indicators. One possible explanation is that because most countries mandate primary education for their citizen while it is not the case for secondary and tertiary education. That being said, changes in primary enrollment ratio does not really affect the education outcomes directly, just like what the biplots show that it does not go to the same direction with other education indicators. However, it is important to note that significant drop in primary enrollment rate might affect secondary and tertiary enrollment rate and therefore it may have an indirect effect on the outcomes of education.

In terms of data summarization, PCA can explain more of the total variance but also produces more outliers, all of them are either poor countries or rich countries with extreme low or high values in two or more variables. While FA cannot explain as much variance as PCA using the same number of factors, it produces only one outlier. Similar case applies to education indicators where FA can explain 5% less variance than PCA, but with the use of varimax rotation, it has the advantage of having only one outlier, compared to 11 outliers with PCA, and the inclusion of significant countries such as China. Considering China is accounted for roughly 20% of world population, it is not a wise decision
to simply exclude it from the dataset. Therefore, in both occasions, FA is more preferable than PCA to summarize this dataset. As for ICT indicators, PCA wins by a margin since it can summarize more variance with fewer outlier than FA.

![Figure 1. PCA biplot for a) all indicators, b) ICT indicators, c) education indicators.](image)

The findings also confirm that digital divide and participation gap do exist to date since poor countries and rich countries tend to fall under different clusters no matter what clustering method is used to the dataset. The dendrogram also identified some countries with extreme values in one or more variables such as United States that produces the most journal articles worldwide and Hong Kong that has the highest mobile subscription worldwide. Both countries lead by a very large margin on each variable respectively, making them very distinctive. The PCA biplots (Figure 2) also show that in poor countries, income is associated more with participation in education while in rich countries it is associated more with the Internet (i.e. Internet users, broadband subscriptions, and Internet servers). It is also interesting to see that in poor countries, broadband and mobile subscriptions tend to be a substitute since they are fetching to the opposite direction from each other. Meanwhile, if we measure
the outcome of education by using the number of journal articles published by scholars of a country, it is associated with the participation in secondary and tertiary education, but only for rich countries. It seems to have nothing to do with the participation in education for poor countries. Again, this confirms that digital divide and participation gap is associated with each other.

Finally, the OLS regression results in Table 1 show that the Internet does promote literacy. It confirms the theory still holds true even after two decades. Using stepwise deletion method by putting all ICT indicators and GDP per capita as independent variables, we can see that for every one more person uses the Internet per 100 people and by holding GDP per capita constant, the literacy rate tends to increase by 23%. Furthermore, when we measure the outcomes of education based on the number of journal and scientific articles published by a country, it tends to increase by 22% for every one additional broadband Internet subscription per 100 people, holding all other predictors constant.

Table 1. OLS regression estimates for education outcomes

|                       | Literacy Rate (^) | Journal Articles (^) |
|-----------------------|-------------------|----------------------|
| GDP per capita        | -.061 ***         | -                    |
| Secure Internet Server| -                 | -1.4 x 10^-3 ***     |
| Internet Users        | .23 ***           | -                    |
| Broadband Subscribers | -                 | .22 ***              |
| Mobile Subscribers    | -                 | -                    |
| Constant              | 4.26 ***          | 3.44 ***             |

Adjusted R²: .655
Highest VIF: 4.31
Number of Cases: 119

Note: Unstandardized coefficient, two tailed-test
* p < .05 ** p < .01 *** p < .001 (^) log transformed

Figure 2. PCA biplot for all indicators in poor countries (left) and rich countries (right).
Furthermore, as shown in Table 2, secondary and tertiary education are indeed highly significant predictors for technology utilization. The higher secondary and/or tertiary enrollment ratio in a country, the more people in that country use ICT. It is interesting to see that secondary education plays more important role than tertiary education for mobile devices penetration. It suggests that the use of mobile devices is highly associated with adolescents more than with adults, just like what was suggested in the literature [8]. It is also worth to note that many researchers found excessive and problematic use of mobile devices among adolescents [9, 10]. Meanwhile, tertiary education is highly associated with broadband Internet connections more than any other education level. This can be explained by the argument that higher education utilizes broadband Internet connection much more than in any other education level through e-Learning platforms and to support research activities as described by the literatures [11]. Considering this kind of activities require large bandwidth and stable connection, it is understandable why it is related more with broadband Internet connection rather than any kind of Internet connection.

Table 2. OLS regression estimates for technology utilization

| GDP per capita  | Secure Internet Server (^) | Internet Users (^) | Broadband Subscriptions (^) | Mobile Subscriptions (^) |
|----------------|----------------------------|-------------------|----------------------------|-------------------------|
|                | 5.33 x 10^{-5} ***        | 5.89 x 10^{-4} ***| 3.21 x 10^{-4} ***        | .10 ***                 |
| Secondary Enrollment | 3.52 x 10^{-2} ***   | .34 ***           | -                          | .41 ***                 |
| Tertiary Enrollment     | 2.47 x 10^{-2} ***   | .31 ***           | .21 ***                    | -                       |
| Constant               | -1.31 **              | -7.41 *           | -1.76 *                    | 1.88 ***                |

Adjusted R^2: .743 .849 .830 .487

Highest VIF: 2.52 2.52 1.46 2.55

Number of Cases: 119 119 119 119

Note: Unstandardized coefficient, two tailed-test
* p < .05 ** p < .01 *** p < .001 (^) log transformed

The results of OLS regression also support the idea of reverse causality between ICT and education. For example, higher tertiary enrollment rate helps promote more broadband Internet subscriptions in a country. On the other hand, having higher broadband Internet subscriptions rate helps foster the number of journal and scientific articles published by the scholars in the respective countries. In other words, we can see the pattern that participation in education promotes more technology utilization, while technology utilization helps improve the outcomes of education.

5. Conclusion

In this paper, several exploration methods are used to confirm that most ICT and education indicators of a country are highly associated with each other as well as with income of the respective country. It is also confirmed that the digital divide in ICT access and participation gap in education are still around and remain as a worldwide problem to solve. Moreover, these two problems are highly associated to each other with income playing a big role in it. The data also suggest that in poor countries, income is associated more with participation in education while in rich countries it is associated more with technology utilization. The findings also support the idea of reverse causality in ICT and education; higher participation rate, especially in secondary and tertiary education level, increases technology utilization while higher technology utilization of a country helps promote better outcomes of education. Combined with the previous findings when income is used as a control variable, it means that people in poor countries are still one step behind when it comes to betterment of education outcomes. Furthermore, it is also interesting to note that fix broadband subscriptions are more associated with tertiary enrollment while mobile subscriptions are more associated with
secondary enrolment. It indicates some differences in needs and/or preferences between adults and adolescent. Finally, both FA, especially with varimax rotation, and PCA have their own benefit and drawback when used to summarize this cross-country dataset consolidated from the 2011-2014 World Bank and UNESCO data. Careful consideration must be taken when choosing either one as a data reduction technique for this dataset.

References
[1] Kuntz P S 1995 Swahili-L: The Internet promotes literacy (Computers & Education) 243 177-81.
[2] Kelley-Salinas G 2000 Different educational inequalities: ICT an option to close the gaps (Centre for Educational Research and Innovation - National Center on Adult Literacy Learning to bridge the digital divide Education and skills)
[3] Ally M (Ed.) 2009 Mobile learning: Transforming the delivery of education and training (Athabasca University Press)
[4] Goksu I and Atici B 2013 Need for Mobile Learning: Technologies and Opportunities. (Procedia-Social and Behavioural Sciences) 103 685-69.
[5] Pick J B and Azari R 2011 A Global Model of Technological Utilization Based on Governmental Business-Investment Social and Economic Factors (Journal of Management Information Systems) 281 49-84.
[6] James J 2012 The ICT Development Index and the digital divide: How are they related? (Technological Forecasting and Social Change) 793 587-94
[7] Pick J B and Nishida T 2015 Digital divides in the world and its regions: A spatial and multivariate analysis of technological utilization (Technological Forecasting and Social Change) 91 1-17
[8] Pratama A R and Al-Shaikh M 2012 Relation and Growth of Internet Penetration Rate with Human Development Level from 2000 to 2010 (Communications of the IBIMA).
[9] Katz J E (Ed.) 2002 Machines that become us: The social context of personal communication technology (Transaction Publishers)
[10] Ha J H, Chin B, Park D H, Ryu S H and Yu J 2008 Characteristics of excessive cellular phone use in Korean adolescents CyberPsychology & Behavior 11(6) 783-4.
[11] Chóliz M 2012 Mobile-phone addiction in adolescence: the test of mobile phone dependence (TMD) Prog Health Sci 2(1) 33-44.
[12] Conole G and Alevizou P 2010 A literature reviews of the use of Web 2.0 tools in Higher Education. (A report commissioned by the Higher Education Academy)