Students' Digital Equality and Scholarly Outcomes

Carson Lam Kai Shun[*]

Alumni of the Faculty of Science, Department of Mathematics, University of Hong Kong, Pok Fu Lam, Hong Kong

Abstract		**Review Article**

Previous work by Lam [1] and Lam [2] show that social objects, cultural identities, and personality form another perspective of Roger Penrose's three world philosophy. Using statistical data from the Centre for Economic and Policy Research (CEPR) in the United States, the relationship between these fields can be determined. This can be achieved through the Principal Component Analysis method by using statistical software: 'R'. It is hoped that this data can be used to validate the sociological aspect of the three worlds philosophy. Certainly, there have been some opposing comments against the proposed theory. As such, this study will investigate some of the main oppositions and suggest possible solutions. Furthermore, by analysing the data, one can study ICT usage amongst students, in addition to predicting their future tendencies. One may explore how the ICT usage may have effects in one's personality and cultural identity together with social objects. Social, cultural, and psychological factors all have effects on the academic achievements of students. These three areas influence the ICT usage of students and their scholarly results. Indeed, the culture and attitude of parents towards their child's ICT usage can affect their educational performance [1]. As a result, the following separate elements arise: 1) Sociological three world philosophy. 2) Parental influence. 3) ICT leadership at school. 4) Studious outcomes. This will finally coalesce into a new and more rationalized connection, which can be presented in a framework. This helps to depict the creative relationship and understand how to minimize digital equality (or inappropriate digital usage) to maximize students' academic performance. Realization can be achieved by manipulating both the three worlds philosophy and parental influence factors. Moreover, a school's successful ICT pedagogy is related to leadership, which is also connected to scholarly results. This means that digital usage can be optimized (or balanced) to produce studious outcomes. As a result, professionals such as teachers, social workers and researchers can develop corresponding strategies, such as providing philosophical education for parents, to better handle the ICT usage of children. Consequently, the digital divide in education, which has been created by modern technology, will be solved.

Keywords: cultural identities, ICT pedagogy, philosophy, education.

INTRODUCTION

How is academic achievement related to educational technology? Previous work by Lam, 03, May, 2016 shows that it is associated with the parents' behavior in guiding their children's ICT usage. As such, there needs to be cultural change among parents as indicated in Lam, October, 2016. Furthermore, it is expected that social objects, cultural identities and personalities influence student' ICT usage. A person's personality is connected to their scholarly results. Therefore, it is interesting to discover the relationship between social, cultural and psychological factors among student ICT usage. The present study will employ statistical data from the Centre for Economic and Policy Research (CEPR) in United States to test as well as principal component analysis using software "R" to test these connections.

LITERATURE REVIEW

Principal Component Analysis (PCA) is used to extract significant components (variables) from the multi-variates among a data set [3]. This allows of the features in a low dimensional set which come from a high dimensional data set. In doing so, this will help to capture as much information as possible. Indeed, lesser variables (principal components or predictors) will encourage meaningful visualization. Thus, PCA is recommended for three or higher dimensional data. The conditions for PCA to work depending on the monotonic relationship between data sets. PCA gives perfect results when the variables are linearly related with one another. However, such a case hardly occurs

naturally [4]. On occasion, one variable will benefit the other one - this is known as axis reflection [5] and suggests an abnormal outcome. There are strategies to compensate for such deficiency as articulated by Mehlman in 1995.

Furthermore, the process of calculating in PCA result is complicated as such computers are usually employed. To further understand the background, the following will explain the basic mathematical theories behind PCA.

Basic Mathematical Theory of Principle Component Analysis

Theoretically, PCA is established on the linear combination of normalized (original) predictors among a data set. Suppose we have a set of predictors (variables) named X^1, X^2, X^3, X^p respectively, they are reduced to two principal components (eigenvectors), say PC_1 (Z) and PC_2(Z), then

$$Z^1 = A^{11}X^1 + A^{21}X^2 + \ldots + A^{P1}X^P$$

Where Z^1 is the first principal component,

A^{p1} is the loading vector. It consists of loadings like $(a^1, a^2, \ldots, a^n)$ for the first principal component. In addition, $a^1 + a^2 + \ldots + a^n = 1$. We note that:

1. A larger magnitude of loadings will result in larger variance. Furthermore, the magnitude will also determine the direction of the data of the principal component (Z^1) along varies.
2. The consequence is then a line with p dimensional space and is closest to no observations. The closeness is defined as the average squared Euclidean space.
3. X^1, X^2, X^p are normalized predictors where the mean equals zero and Standard derivation equals 1.

From the above, one observes that the first principal component (Z^1) can be expressed as a linear combination of the initial variables. Through it, the maximum variance of the data set can be held. The most feasible direction for the variability of the data set can also be determined. Indeed, the component can capture more information when the variability is larger. Finally, the resultant straight line means the component can be used to minimize the sum of the squared distance between a data point and the line.

In a similar way, the second principal component (Z^2) can also be expressed as the linear combination of the original variables. The component holds the outstanding part of variance in the data set and is uncorrelated to Z^1. The correlation between Z^1 and Z^2 is zero, hence their directions should be orthogonal. Practically, the second principal component is represented as:

$$Z^2 = A^{12}X^1 + A^{22}X^2 + \ldots + A^{P2}X^P.$$
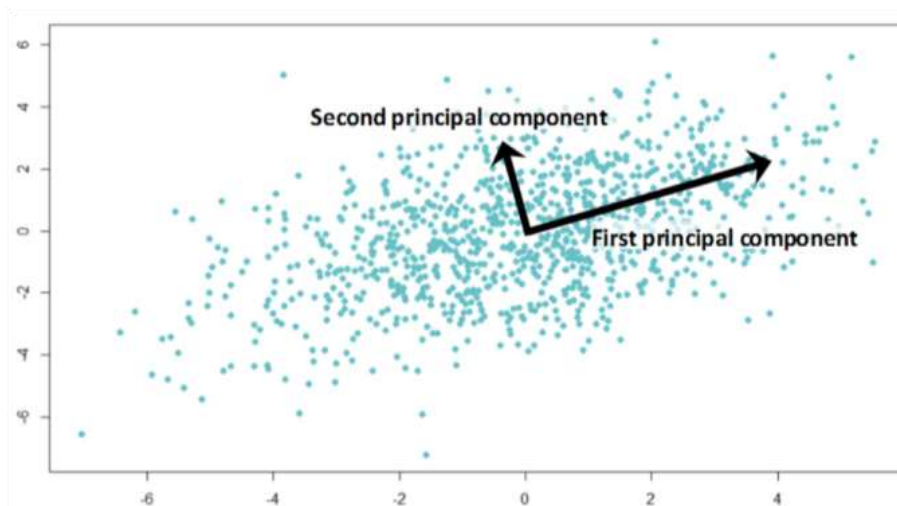
The data can be displayed as below:

Fig-1: Uncorrelated data[2] between Z^1 and Z^2

To conclude, one can apply a similar concept to the other principal components. In general, given any n x p dimensional data, one can construct a minimum of $(n-1)$ p principal components. However, the direction of these components is unsupervised and cannot be determined by the dependent variable. It is because the direction can only be judged by eigenvectors calculated from the covariance matrix. This is proposed that Partial Least Square (PLS) should be used. PLS assigns high weight to those variables which are strongly related to dependent variables and thus solves the problem.

Furthermore, the original predictors should be normalized as different variables may use different scale. Un-normalized PCA will lead to large loadings of components. Hence the dependence of an eigenvector on the high variance predictor which is undesirable.

Practical Steps in performing Principal Component Analysis [6]
Step 1: Calculate the mean of observed (input) data X.

When measuring a single variable A. Let the n measurements be denoted as a1, a2,…, an, then the mean (sample average) μA is given by:

$$\mu_A = \frac{1}{n} a_1 + a_2 + \cdots + a_n$$

Hence, the measurements are centered [7]. The spread out of the measurements is then given by:

$$\text{Var(A)} = \frac{1}{n-1}[(a_1 - \mu_A)^2 + (a_2 - \mu_A)^2 + \ldots + (a_n - \mu_A)^2]$$

The covariance of the two variables A and B is given by:

$$\text{Cov(A,B)} = \frac{1}{n-1}[(a_1 - \mu_A)(b_1 - \mu_B) + (a_2 - \mu_A)(b_2 - \mu_B) + \ldots + (a_n - \mu_A)(b_n - \mu_B)]$$

When calculating the mean of m variables, then μ is a single vector in $\mathbf{R^m}$
In general:

$$\boldsymbol{\mu} = \frac{1}{n}(X|\vdots|1 + X_2 + \cdots + X_n)$$ where X_i is a sample vector.

Example 1:

$$X_1 = \begin{pmatrix} 2.4 \\ 2.5 \end{pmatrix}, X_2 = \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix}, X_3 = \begin{pmatrix} 2.2 \\ 2.9 \end{pmatrix}, X_4 = \begin{pmatrix} 1.9 \\ 2.2 \end{pmatrix}, X_5 = \begin{pmatrix} 2.4 \\ 2.5 \end{pmatrix}$$

$$X_6 = \begin{pmatrix} 2.3 \\ 2.7 \end{pmatrix}, X_7 = \begin{pmatrix} 2.0 \\ 1.6 \end{pmatrix}, X_8 = \begin{pmatrix} 1.0 \\ 1.1 \end{pmatrix}, X_9 = \begin{pmatrix} 1.5 \\ 1.6 \end{pmatrix}, X_{10} = \begin{pmatrix} 1.2 \\ 0.9 \end{pmatrix}$$

Then the mean $\mu_A = \frac{1}{10} = 1.81$;

$$\mu_B = \frac{1}{10}(2.4 + 0.7 + 2.9 + 2.2 + 3.0 + 2.7 + 1.6 + 1.1 + 1.6 + 0.9) = 1.91.$$

Step 2: Calculate the covariance matrix S of the input data X.
To "center" the data, the mean needs to be subtracted from each sample vector Xi. Let B be the m x n matrix, then

B = [X1 – μ, … , Xn - μ]

Next, the covariance matrix S is defined as $S = \frac{1}{n-1}BB^T$.

S is symmetric. The following two theorems can be noted:

Theorem 2: If A is symmetric (i.e. $A^T = A$), then A is said to be orthogonally diagonalizable and has only real eigenvalues. In other words:

$$Ax_i = \lambda_i x_i$$

Where $\lambda_1, \lambda_2, \ldots, \lambda_n$ (named as eigenvalues) are real numbers
X1, X2, …, Xn (named as eigenvectors) are non-zero real orthogonal vectors

Corollary 3: If A is any m x n matrix of real numbers, then both the m x m matrix AA^T and the n x n matrix A^TA are symmetric.

Furthermore, suppose

$$X_1 = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}, X_2 = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}, X_3 = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}$$

then we get, $$\begin{pmatrix} a_1 - \mu_1 & b_1 - \mu_1 & c_1 - \mu_1 \\ a_2 - \mu_2 & b_2 - \mu_2 & c_2 - \mu_2 \\ a_3 - \mu_3 & b_2 - \mu_2 & c_1 - \mu_1 \\ a_4 - \mu_4 & b_4 - \mu_4 & c_4 - \mu_4 \end{pmatrix}$$

Consider the 1, 1 entry of S, we have

$$= S_{11} = \frac{1}{3-1}[(a_1 - \mu_1)^2 + (b_1 - \mu_1)^2 + (c_1 - \mu_1)^2]$$

Which is the variance of the first variable.
Furthermore, consider the 2,1 entry of S:

$$S21 = \frac{1}{3-1}[(a_1 - \mu_1)(a_2 - \mu_2) + (b_1 - \mu_1)(b_2 - \mu_2) + (c_1 - \mu_1)(c_2 - \mu_2)]$$

In general, for $1 < i, j < m$:
Sii is the variance of the i th variable and Sij is the covariance between the i th and j th variables.

Example 1 (con't): After lengthy calculation (follow the step 2), we get the covariance matrix as:

$$Cov\ S = \begin{pmatrix} 0.61655556 & 0.615444444 \\ 0.615444444 & 071655556 \end{pmatrix}$$

Step 3: Calculate the eigenvalues and eigenvectors of the covariance matrix.

Consider the following eigenvalue problem which have the format:

$$\mathbf{A}.\ \mathbf{v} = \lambda.\ \mathbf{V}$$

where $\mathbf{A}$ is m x m matrix, $\mathbf{v}$ is m x 1 non-zero vector,
λ is a scalar

Any value of λ which satisfies the above equation has a solution is termed as the eigenvalue of $\mathbf{A}$

while the corresponding vector $\mathbf{v}$ is called the eigenvector of $\mathbf{A}$.

By using simple matrix algebra:

$$\mathbf{A}.\ \mathbf{v} = \lambda.\ \mathbf{v}$$
$$\mathbf{A}.\ \mathbf{v} - \lambda.\ \mathbf{I}.\ \mathbf{v} = \mathbf{0}$$
$$(\mathbf{A} - \lambda.\ \mathbf{I}).\ \mathbf{v} = \mathbf{0}$$

Then, the roots of $|\mathbf{A} - \lambda.\ \mathbf{I}|$ will gives the eigenvalues and the corresponding eigenvectors.

Example 1 (continue):

$$|\mathbf{A} - \lambda.\ \mathbf{I}| = |\begin{pmatrix} 0.61655556 & 0.615444444 \\ 0.615444444 & 071655556 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}|$$

After calculation, $\lambda1 = 0.490833989$ or $\lambda2 = 1.28402771$

The corresponding eigenvectors are: $\begin{pmatrix} -0.735178656 \\ 0.677873399 \end{pmatrix}, \begin{pmatrix} -0.677873399 \\ -0.735178656 \end{pmatrix}$

Step 4: Computing the feature vector (matrix) and the principal components Pi. Example 1 (continue): The corresponding feature vector (Matrix) Fv is given by: Fv = (eig1, eig2, , eign)
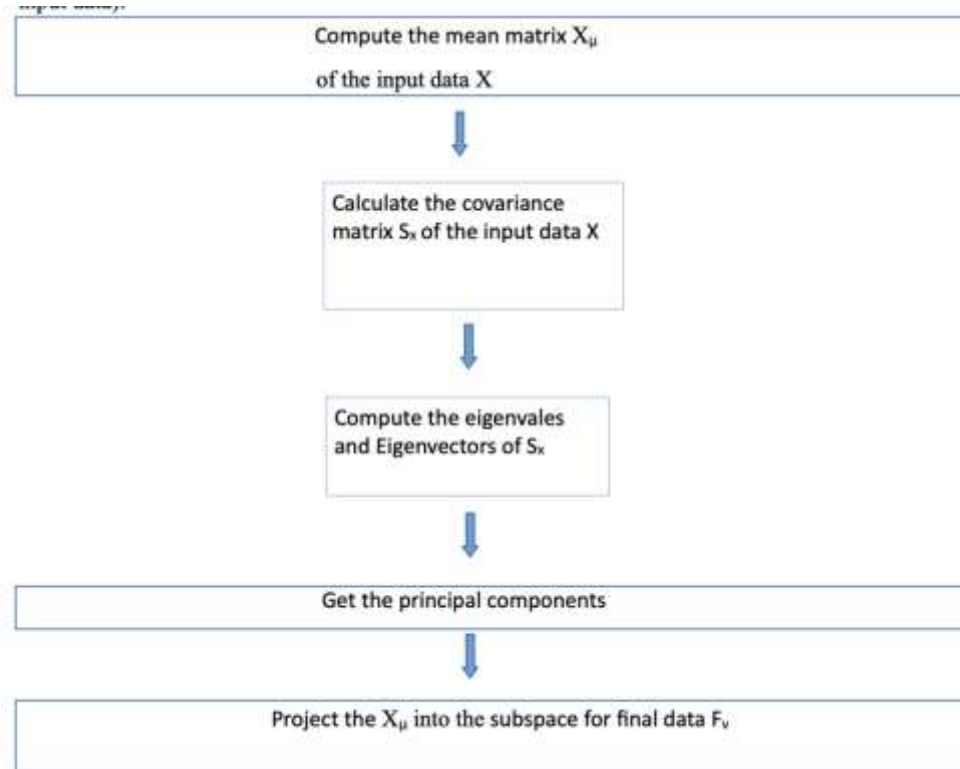
$$F_v = \begin{pmatrix} -0.735178656 & -0.6778773399 \\ 0.6778773399 & -0.735178656 \end{pmatrix}$$ or P_1 = -0.74X + 0.68Y, P_2 = -0.68X + (-0.74)Y

Step 5: Calculating the final data. Example 1 (con't)

$$\underline{T \quad T}$$

FinalData FD= RowFeatureVector FRVx RowZeroMeanData FZMD where FRV is the transpose of Fv FZMD is the mean-adjusted data transposed

$$\text{i.e. } \begin{pmatrix} -0.735178656 & -0.6778773399 \\ 0.6778773399 & -0.735178656 \end{pmatrix} \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ -0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix} = \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ -0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix}$$

Compute the mean matrix X_μ
of the input data X

Calculate the covariance matrix S_x of the input data X

Compute the eigenvales and Eigenvectors of S_x

Get the principal components

Project the X_μ into the subspace for final data F_v

To summarise, the following algorithm, in form of a flowchart, can be used to calculate the principal components and final data [8] when given a transformation Y = PX (X is the input data):

Some Interpretations of the Principal Components Analysis
Once the principal components of the original data X have been obtained, they then be interpreted. The following sections will discuss.

Finding the Eigenvectors of Covariance for (the transformed) matrix Y
Let **X** and **Y** be two m x n matrices which are related by a linear transformation **P** [9]. **X** is the data set, m is the number of the measurement types and n is the number of data type trials.

i.e. An orthonormal matrix **P** is desired such that:

$\mathbf{Y} = \mathbf{PX}$ and $\mathbf{SY} = \frac{1}{n-1} \mathbf{YY}\mathrm{T}$ is diagonalized, **SY** is the covariance matrix of **Y**.

Furthermore, the rows of P are the principal components of **X**.
Consider the covariance matrix **SY** which is in terms of our choice **P**, then we have:

Theorem 4: A square matrix is symmetric if and only if it is orthogonally diagonalizable. Furthermore, it is diagonalized by a matrix of orthonormal eigenvectors.

$$\mathbf{SY} = \frac{1}{n-1}\mathbf{YY}^{\mathrm{T}}$$
$$= \frac{1}{n-1}(\mathbf{PX})(\mathbf{PX})^{\mathrm{T}}$$
$$= \frac{1}{3-1}\mathbf{P}\ \mathbf{X}\ \mathbf{X}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}$$
$$= \frac{1}{3-1}\mathbf{P}\ (\mathbf{XX}^{\mathrm{T}})\mathbf{P}^{\mathrm{T}}$$
$$\mathbf{SY} = \frac{1}{3-1}\mathbf{P}\ \mathbf{A}\ \mathbf{P}^{\mathrm{T}}$$

Where $\mathbf{A}$ is a newly defined matrix such that $\mathbf{A} = \mathbf{XX}^T$ and $\mathbf{A}$ is symmetric

Hence, $\mathbf{A} = \mathbf{EDE}^T$, where $\mathbf{D}$ is a diagonal matrix
$\mathbf{E}$ is a matrix of eigenvectors of $\mathbf{A}$ arranged in columns.
The matrix A has a dimension $r < m$ where r is the number of orthonormal vectors (which will be discussed in the next section).

With a suitable selection of matrix $\mathbf{P}$ which has each row $\mathbf{P}i$ as an eigen-vector of
$\mathbf{XX}^T$, $\mathbf{P} = \mathbf{E}^T$. This implies that
$\mathbf{A} = \mathbf{P}^T\mathbf{DP}$

Theorem 5: The inverse of an orthogonal matrix equals to its transpose.
From theorem 5, $\mathbf{P}^T = \mathbf{P}^{-1}$, therefore we have:
$\mathbf{A} = \mathbf{P}^{-1}\mathbf{DP}$ or
$\mathbf{SY} = \frac{1}{n-1} \mathbf{P}\,\mathbf{A}\,\mathbf{P}^T$

$$= \frac{1}{n-1}\,\mathbf{P}(\mathbf{PTDP})\mathbf{PT}$$
$$= \frac{1}{n-1}\,(\mathbf{PPT})\,\mathbf{D}\,(\mathbf{PPT})$$
$$= \frac{1}{n-1}\,(\mathbf{PP}\text{-}1)\,\mathbf{D}\,(\mathbf{PP}^{-1})$$
$$\mathbf{SY} = \frac{1}{n-1}\,\mathbf{D}$$

The above steps explain how to diagonalize $\mathbf{SY}$ which is the ultimate goal.

Dimension Reduction in Principal Component Analysis
In the following section, we assume M is a real square symmetric matrix[3], dimension r x r

With eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ and the corresponding orthonormal eigenvectors u1, u2,………….. , ur.

Theorem 4 suggests,

$$1.\ M = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_r \end{bmatrix} \begin{bmatrix} \leftarrow u_1 \rightarrow \\ \leftarrow u_2 \rightarrow \\ \leftarrow u_r \rightarrow \end{bmatrix}$$
named as Q Λ Q^T

$$2.\ M = \sum_{i=1}^{r} \lambda_i u_i u_i^T$$

$z \in \mathbb{R}^r$ Definition 6: With the assumption of M at the beginning of this section, if $z^T M z \geq 0$ for all
$z \in \mathbb{R}^r$ If $z^T M z > 0$ for all nonzero, it is called a positive definite.

$X \in \mathbb{R}^r$ For instance, given a random vector, let $\mu = EX$ and $S = E[(X - \mu)(X - \mu)^T]$ its mean and covariance.
$z^T S z = z^T E[(X - \mu)(X - \mu)^T]z = E[z^T(X - \mu)(X - \mu)^T z]$
$= E[(z.(X - \mu))^2]$
≥ 0

Obviously, S is a positive semidefinite.

Theorem 7: If M is a real r x r symmetric matrix, then
1. M is a positive semidefinite only if it has all eigen-values $\lambda i \geq 0$.
2. M is a positive definite only if it has all eigenvalues $\lambda i > 0$.

Theorem 8: With the assumptions, we have the following:

$$\min_{\|z\|=1} z^T M z = \min_{z \neq 1} \frac{z^T M z}{z^T z} = \lambda_r$$

Are realized at z = u1 and z = ur, respectively.

For example, consider the random vector $X \in \mathbb{R}^r$ which has mean μ and covariance matrix M, then the variance of X in the direction z is represented by $z^T M z$ such that:

$$\text{var}(z^T X) = E[(z^T(X - \mu))^2] = E[z^T(X - \mu)(X - \mu)^T z] = z^T M z$$

Theorem 8 implies that the variance will attain its maximum value in the direction of u1 and its minimum in the direction ur.

Then one may wonder which r-dimensional $\mathbb{R}^r$ subspace will have the most $\mathbb{R}^k$ variance. In fact, for a linear projection from to, a function $x \mapsto P^T x$,

Where P^T is a k x r matrix with $P^T P = \mathbf{I} k$.

This shows that the rows of projection matrix are orthonormal. Moreover, the covariance matrix of

$$\max_{\|z\|=1} z^T M z = \max_{z \neq 1} \frac{z^T M z}{z^T z} = \lambda_1$$

the resulting k-dimension vector is: $cov(P^TX) = E[P^T(X − \mu)(X − \mu)^TP] = P^TMP$.

In order to represent the entire matrix by a single number, one may express it in terms of trace i.e. $var(P^TX) = tr(P^TMP) = E\| P^TX − P^T\mu\|$. Indeed, it is desirable to maximize the variance of projection P^T so that the signal to noise ratio (SNR) is much greater than one. The higher the SNR value, the higher the precision of data [9]. Conversely, a low SNR signifies that there is noise contaminating the data. Thus, the following theorems are useful in determining (or optimizing) the maximum value of variance:

Theorem 9: Let M be a real r x r symmetric matrix as in theorem 7. Pick any $k \leqslant r$.

$$\max_{P\in R^{rXk}, PP^T=I} tr(P^TMP) = \lambda_1 + \lambda_2 + \ldots + \lambda_k$$
$$\min_{P\in R^{rXk}, PP^T=I} tr(P^TMP) = \lambda_{r-k+1} + \lambda_{r-k+2} + \ldots + \lambda_r$$

These are realized when the columns of P span the k-dimensional subspace spanned by {u1, u2, …, uk} and {ur-k+1,…, ur} respectively.

Proof:
Only the case for maximum should be proved. Suppose p1, p2, …, pk are the

Columns of P, then

$tr(P^TMP) = \sum_{i=1}^{k} P_i^T MP_i = \sum_{i=1}^{k} p_i^T \left(\sum_{j=1}^{r} \lambda_j u_j u_j^T\right)p_i$
$= \sum_{i=1}^{r} \lambda_j \sum_{j=1}^{k}(p_i u_j)^2$

Let z_j be $\sum_{j=1}^{k}(p_i u_j)^2$ which is non-negative, we want to show $\sum_{j=1}^{r}(z_j) = k$ and that $z_j \leq 1$.

Initially,
$\sum_{j=1}^{r}(z_j) = \sum_{i=1}^{k}\sum_{j=1}^{r}(p_i u_j)^2 = \sum_{i=1}^{k}\sum_{j=1}^{r}(p_i^T u_j u_j^T p_i)^2$
$= \sum_{i=1}^{k}\sum_{j=1}^{r}(p_i^T QQ^T p_i)^2 = \sum_{i=1}^{k}\|P\|^2 = K$

To find the upper bound of z_j, the k orthonormal vectors $p_1, p_2, …, p_k$ can be extended to a full orthonormal basis $p_1, p_2, …, p_r$ of $\mathbf{R}^r$. Thus, we have:
$z_j = \sum_{j=1}^{k}(p_i u_j)^2 \leq \sum_{j=1}^{r}(p_i u_j)^2 = \sum_{j=1}^{r}(u_j^T p_i p_i^T u_j)^2 = \|u_j^T\|^2 = 1$

Then we conclude that,
$tr (P^TMP) = \sum_{j=1}^{r} \lambda_j z_j \leq \lambda_1 + … + \lambda_r \leq \lambda_1 + … + \lambda_k$

We remark that since $\sum_{j=1}^{r}(z_j) = k$ (where k < r) with $z_j = 1$, thus there must be r-k+1 number of zeros.

Hence, the required result follows. The equality holds where $p_1,…,p_k$ span the same space as $u_1, u_2,…, u_k$.

Let, we want to find such that $\|p\| = 1$ and maximize var(pTX).

$$X \in \mathbb{R}^r \qquad p \in \mathbb{R}^r$$

Theorem 10: To optimize (max var(pTX)) the above problem, p should be the principal eigenvector (corresponding to the largest magnitude of eigenvalue) i.e p1 of cov(X).

Concerning the k-dimensional subspace, when will it be the best approximation to X?

Theorem 11: Consider the mean (μ) and covariance (S) of X, the way to minimize the error (best approximation) to X is the choosing of the top k eigenvectors say p1, p2,………., pk of S and set p0 = (I − PT)μ

Proof: Fix any given matrix P; one need to choose a p0 that will minimize $E\|X − PTX + P0\|2$. This is calculated as:
$\frac{d}{dx}\{E [X − (P^TX + P_0)]^2 = 2E [X − (P^TX + P_0)] (1-P^T) = 0$
$E [X − (P^TX + P_0)] = 0$
$E [X (1-P^T)] − E[P_0] = 0$
$E[P_0] = E (1-P^T) E (X)$
$P_0 = (1-P^T)$

To optimize P, the cost function is:
$E\|X − (PTX + P0) \|2 = E\|(I-PT)(X-\mu)\|2 = E\| X-\mu\|2 − E\|PT(X-\mu)\|2$

Where the second step is only the Pythagorean theorem.

Theorem 12: For the random vectors X and Y, $E[\|X − Y\|2 - \|PTX − PTY\|2]$ is minimized when the rows of PT are top k eigenvectors of cov(X). Note: PT is a k x r projection matrix.
Theorem 13: Cost ($\mu1, …..,\mu k$) $\leqq$ 2*OPT.

Lemma 14: MTM and MMT are symmetric positive definite matrices.

Lemma 15: If λ is an eigenvalue of MTM with eigenvalue u, then either λ is an eigenvalue of MMT with eigenvector Mu, or $\lambda = 0$ and Mu = 0.

Theorem 16: Let M be a rectangular m x n matrix with m $\leqq$ n. Define λi , ui, vi as the eigenvalues and the eigenvectors of (larger) matrix M^TM. Also, vi be the eigenvectors of MM^T. Then:

$$1. \ M = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_m \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_m} \end{pmatrix} \begin{pmatrix} \leftarrow u_1 \rightarrow \\ \leftarrow u_2 \rightarrow \\ \vdots \\ \leftarrow u_r \rightarrow \end{pmatrix}$$

Q₁, size m x m   Σ, size m x n   Q₂ᵀ, size n x n

To conclude, from the aforementioned theorems and proofs (for the case of square matrix which can be extended to a rectangular one), one may discover why (i.e. tries to increase the SNR that results in high precision data) and how to select k vectors which span the k-dimensional subspace for capturing as much as possible of variance of X (i.e. chooses the top k eigenvectors of the cov(X) - S). Thus it is reasonable for us to reduce the dimension of the random variable X through the procedure of principal component analysis and observe the relationships between different principal components.

Models for Principal Components Analysis by Singular Value Decomposition

What is meant by the term model? According to Walter [10], a model can give us a mathematical description of the observing process to a set of data. In particular, a statistical model is an equation which describes the:

1. The impact of explanatory variables and
2. The probability distributions that are characterized by the random variation process.

In other words, statistical models include two parts – a systematic component and a random component. i.e.

$y_{ij} = \mu_i + e_{ij}$

Where i = 1,2 means treatment,

j = 1, 2,…………..ni, means the observation j,

nj means the times of observation in i th treatment, yij means the jth observation occurred in the i th treatment μi determines the i th treatment's mean eii determines the ij th observation's random "error".

Firstly, the systematic part of the model is determined by the treatment it receives, therefore it is also called the deterministic part of the model. This means the first part is a mathematical law without any random variability. Secondly, the random part explains how observations are varied randomly about their mean. Indeed, one can write a completely deterministic model under certain schools of philosophy when one knows what made the ij th individual unique. By unique ij th individual means the term random error eij. However, in this instance we will not characterize the observation distribution by probability. Normally, we assume eij to be N (0, σ2) [10].

1. The probability distributions that are characterized by the random variation process.

In other words, statistical models include two parts – a systematic component and a random component. i.e.

$y_{ij} = \mu_i + e_{ij}$

Where i = 1,2 means treatment,

j = 1, 2,…………..ni, means the observation j,

nj means the times of observation in i th treatment, yij means the jth observation occurred in the i th treatment μi determines the i th treatment's mean eij determines the ij th observation's random "error".

Firstly, the systematic part of the model is determined by the treatment it receives, therefore it is also called the deterministic part of the model. This means the first part is a mathematical law without any random variability. Secondly, the random part explains how observations are varied randomly about their mean. Indeed, one can write a completely deterministic model under certain schools of philosophy when one knows what made the ij th individual unique. By unique ij th individual means the term random error eij. However, in this instance we will not characterize the observation distribution by probability. Normally, we assume eij to be N (0, σ2) [10].

In the case of PCA by singular value decomposition, there are three fundamental models:
1. Descriptive algebraic model [11]

When the model is reduced to its simplest form, this is the SVD. Instead of simply using a least square method, a necessary metric can be chosen instead. This model does not contain any random elements and implies that there is no expectation or variance. It also gives a geometric view of PCA. Most of the descriptive algebraic models were described in the previous section.

2. Random effects model [12]

This is applicable in the case of dataset observations which represent a sample from a very large population. In fact, standard parametric procedures cannot be employed such that the capacity of PCA could be determined. The model is evaluated through a computer-based re-sampling technique called "jackknife". The learning set used for estimating left-out observations is known as testing set. A random

effect model is used and the predictions are stored in a matrix $\tilde{X}$.

3. Fixed effects model [11]
This model the rows x1, x2,……….., xn of X are independent random variables with

$$\frac{\sigma^2}{w_i}$$

facts that E(xi) = zi such that zi is in a q-dimensional subspace, Fq. In addition, if ei = xi − zi, then E(ei) = 0 with var(ei) = Γ. Indeed, Γ is a positive definite symmetric matrix and wi is a positive scalar with a sum of 1.

The major point of interest for the fixed model lays in the variation of observed features among the means of the dataset. In the present study, none of the above can best fit the data since they have both random and fixed effects. It is therefore proposed that one should reassemble the aforementioned models for the PCA by SVD. Then the proposed system can go beyond the current statistical framework of the research. In other words: "The evaluation of models touches on deep issues in the philosophy of science, because the statistical model often determines how the data-generating system under investigation is conceptualized and approached [13]. Model choice thus resembles the choice of a theory, a conceptual scheme, or even of a whole paradigm, and thereby might seem to transcend the formal frameworks for studying theoretical rationality [14, 15, 6]."

To be more precise, we can combine the original three systems and transcend them into a mixed (or linear) model for both the random and fixed effects. It is in fact a joined analysis [16] which shows in the following equations:

$$y = X\beta + Zu + e \ldots\ldots\ldots\ldots \text{ Equation (1)}$$

Where y is the vector of the transformation, X, Z are the incidence matrices for fixed and random effects respectively, e is the random error vector, β and u are the fixed and random effects vectors corresponding with the assumptions that e ~ N(0, R) u ~ N (0, G) **R**, **G** is the covariance matrices of e and u respectively.

Indeed, the equation (1) can be written into a "matrix formed" equation:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & . & . & . & . \\ . & x_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & x_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_n \end{pmatrix} + \begin{pmatrix} z_1 & . & . & . & . \\ . & z_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & z_n \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix}$$

The joint probability density function for y and u is: f(y,u) = f(y|u) f(u). Then according to Henderson's "mixed model equations' MME" [17, 18]:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mu \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

Where β and are μ called the best linear unbiased estimates (BLUE) and predictors (BLUP) for β and u, respectively.

According to the Gauss-Markov theorem, one cannot scale the result of conditional variance to the identity matrix. If one can have idea of conditional variance, then one will find that the inverse variance of weighted least square is BLUE. To solve MMEs, one need to estimate both the variance and weighted parameter since there is a difficulty to obtain the conditional variance. The EM algorithm can be used for fitting the mixed models.

This is shown as follows [16]:
Let a matrix $C = \begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$, then for the matrix G we have:

$$\sigma_{uij} = [u_i^T u_j + tr (C_{ij}^{-1})] / t$$

Where $\sigma_{uij} = \sigma_{uk}^2$ if i =k
σ_{uij} if otherwise

The residual covariance estimator contained in R is given by:

$$\sigma_{eij} = \{e_i^T e_j + tr ([KC^{-1}K])\} / n*$$

Where $\sigma_{eij} = \sigma_{ek}^2$
   if i = k
   σ_{eij} if otherwise

Where K = {X,Z}, the trace depends on the index i and j and n* denotes the length of the vector {i,j}.

In terms of computer programming ideas, the 'lme4' package in software R is employed in this study for fitting the research data. Indeed 'lme4' is designed philosophically opposite to the concept of p-values [19]. In other words, the 'car' packages will not be used which is focused mainly on the null hypothesis −

determining the confident of effects between different variables. What needs to be determined, is the inter-relationship across different variables in the research of the sociological three world philosophy. In other words, (Linear mixed model fit by REML [20]):

$$y \sim x1 + x2 + ... + (1 \mid r1) + (1 \mid r2) +.....$$

Where y is the response variables, x1, x2, … are explanatory

Variables with added in random effects,

$(1 \mid r1) + (1 \mid r2) +…$ are the crossed random effects

i.e. $y \sim (x'1 + v1) + (x'2 + v2) + ...+ (1 \mid r1) + (1 \mid r2) +.....$

where $x'i$ are explanatory variables and $v1$, $v2,…$ are the remaining added in random effects Moreover, $x'i + vi = xi$.

To go into more detail, one can combine the thinking behind a proposed mixed model and sublimate it into a new level of concept which forms a generalized case [21],

$$y \sim (x'1\ \beta + z1v1) + (x'2\ \beta + z2v2) +..... + (1 \mid r1) + (1 \mid r2) +.....$$

Where β is the fixed effects of the explanatory variables xi , zi are the variables with random effects

The above is the singular case of expressing principle component analysis by a linear regression relationship. When we have a different linked case of casual relations that expressed in terms of:

$$Y = XB + B_0 \text{ and } Z = YD + D_0;$$

The aforementioned method is known as partial least squares method for column vector Y and Z or matrices Y and Z. Actually, for a series of domino effect, one will have:

$$Y = XB + B_0$$
$$Z = YD + D_0 \text{ or } Z = \{XB + B_0\} D + D_0$$

This means one can always express the series of domino effect in a sequence of recursive approximated manner or a partial least square regression.

To be precise, the Bayesian Matrix, say [M], can be expressed by the regression as:

$$[M] [LT] = X + (XB + B_0) + \{(XB + B_0)D + D_0\}$$

Where [LT] is the associated linear transformation; while the converse is also true:

$$X + (XB + B_0) + \{(XB + B_0)D + D_0\} = [M] [LT]$$

We may further deduce a nested case of principle component analysis typed linear regression like the situation in Lam March, 2020:

$$(x'_1 + x'_2 +…) + (x'_1\ \beta_1 + x'_2\ \beta +…) + (x'_1\ \beta_1\ \mu_1 + x'_2\ \beta_2\ \mu_2 + …) + (z_1 + z_2 +…) + (z_1 v_1 + z_2 v_2 + …) + (z_1 v_1 \alpha_1 + z_2 v_2 \alpha_2 +…) + …$$

Or we may have the following mathematical expression:

$$(1+B+BD) X + B_0 D + B_0 + D_0 \text{------------------ eqt (*)}$$

If we compare eqt(*) with the matrix equation (step 4) in Baron and Kenny method:

$$Y = B_6 + B_7 X + B_5 M + e3$$

We find that D is in fact a mediator.

When we have applied the above result in the principal component analysis, we may generally get:

$$y \sim (1+\beta_1+\beta_1\mu_1 +...) X_1'' + B_0 D + C_0 + (1 + v_1 + v_2\alpha_2 +...) Z_1'' + E_0 F + G_0 +…$$

Where X_1'' and Z_1'' are the vectors $(x_1', x_2'…)$ and $(z_1', z_2'…)$

In other words, both matrix D and F act as the mediator for X_1'' and Z_1'' correspondingly. Or the variable y is related to the wanted Hayes models in terms of the principal components X_1'' and Z_1''. In fact, we have found the model for the variable y.

RESULTS

After reviewing the method and models in statistics, this section covers the data results. Various outcomes of social, cultural and psychological aspects will be presented to show their relationship between ICT usage and academic achievement.

Social Objects

According to Lau [22], skill disparity in the areas of Internet and computer literacy, computer access ratio and learning related usage influences academic performance. Furthermore, social objects such as socio-economic status, ICT experience and gender have an influence on skill, usage and ethical disparities. Finally, parental ICT abilities, permission, worry, monitoring and encouragement also affected these disparities.

In short, Lau explains us that students' computer and Internet literacies helps them with activities such as submitting assignments, preparing presentation slides and searching for information through Internet etc. Moreover, his research clearly shows learning-related usage has a positive effect on studying but a negative one in the case of leisure-usage. The results are like Lam, 2014 who proposed ICT usage can affect academic performance under the conditions of "positive and quality usage" [23]. In addition to the above outcome, Lau also verifies that parental ICT abilities, permission, worry, monitoring and encouragement all play an important part in affecting the disparities – access, skill, usage and ethics. Finally, social objects such as socio- economic status, ICT experience and gender also have an influence on these

disparities. The main findings and reasons are as follows [22]:

1. Students with low SES (Social Economic Status) usually lack resources and therefore need to share digital devices with other family members. This tells us why SES affects access disparity.
2. Students with high SES are less likely to use digital devices for leisure-related purposes. This is because their parents can afford to pay more expensive activities for their children such as learning a musical instrument or foreign language. As a result, high SES children get a broader life experience. This shows how SES affects usage disparity.
3. Low SES students always use the excuse that a small amount of money does not affect huge software companies. As such, they do not understand the significance of software piracy. This provides a reason for the positive relation between ethical disparity and ethical perception.
4. Parental ICT ability has a positive effect on

student access and skill disparity. Parents with a low ICT ability usually have a low educational background and consequently a low SES. Hence, low SES students are likely to have a disadvantage in access disparity.
5. Parental monitoring partly has an influence on usage disparity but is positively connected with all student learning and leisure-related usage. When parents are closely monitoring their children's activities, they will spend more time at home using ICT regardless of whether for learning or leisure. In front of their parents, students tend to do what they are told. Without supervision, it is less likely they will follow instruction.

In general, parents and social objects have effects to all four disparities as studied by Lau, 2014 [22]. Hence, by following Lau's consequences, this author suggests that one can establish a relation between social objects, parents, ICT usage and academic performance as depicted in the next page.

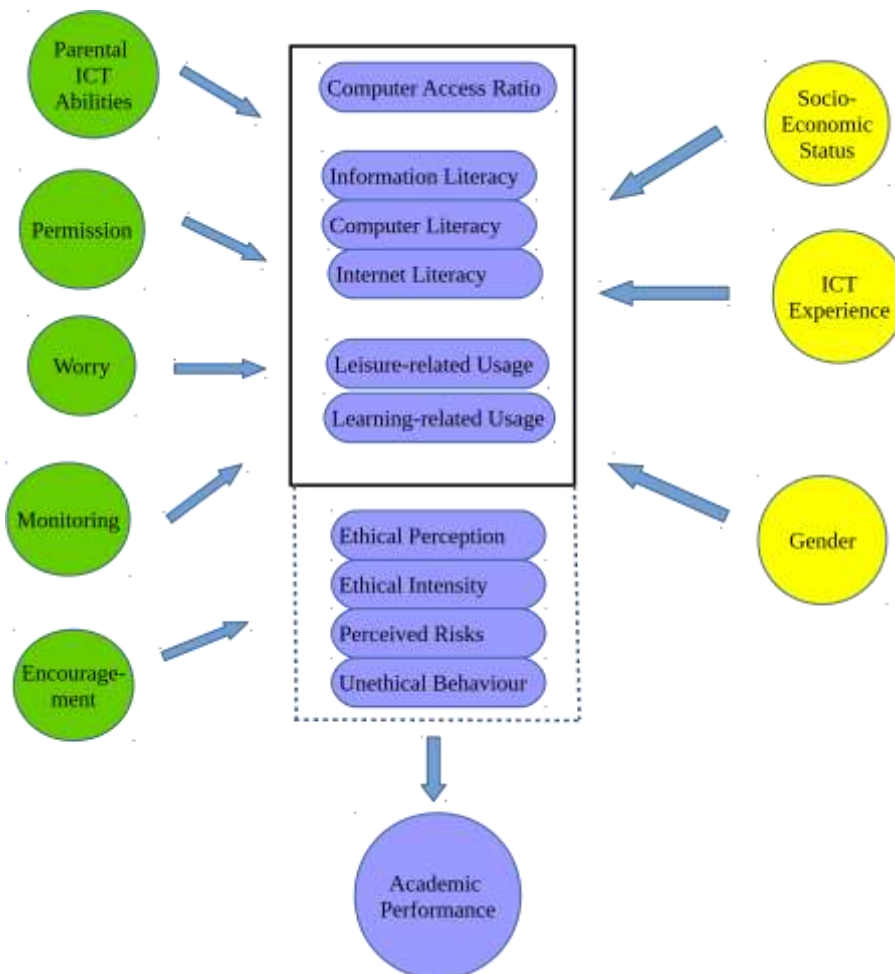**Fig-2: The relationship between social objects, students' ICT usage, parental influence and academic performance**

Besides the above connection, one can also ask whether there is any relation between social objects and cultural identities? In answer to that question this study

presents statistical data showing the social, cultural and psychological factors of students accessing digital tools.

The data was obtained from the Centre for Economic and Policy Research (CEPR) in United States

Correlation between social objects, cultural identities and personality

This section shows the results by using a linear mixed model and analyses the relationship between social objects (gender, age, income, family, education, race, region and state) and cultural identities (English, language spoken at home, main language used, limited English speaking in household and main household language) in the areas of accessing digital tools and the Internet (Internet access, broadband and laptop). One of the main cultural identities for student is known as language. For examples, in the late 80s and early 90s, most local Chinese people in Hong Kong spoke Cantonese supplemented by a few words in English. Foreigners would recognize this and immediately identify them as 'Hong Kongers', very different from Mainland Chinese. The difference being that while the locals were Chinese and spoke Cantonese but they lived in Hong Kong under British rule – where the official language was English. This author suggests that it is this unique background which creates Hong Konger's cultural identity.

Data Processing Procedure

To begin with, data was taken from the website of the Centre for Economic and Policy Research (CEPR). Statistical software such as 'Wizard', was then used to select information on information under the age of thirty. This was done to reduce the file size since the original data was about 30 million. The resultant file contained around 7.8 million of data and was saved in R format. This was done as is much easier to manipulate. Next, the software program Stat/Transfer was used to convert the file into SPSS plus ACSII format. The reason for this was to automatically convert the data from text to numeric format. As a result, it could easily be handled with R's Generalized Linear Mixed Model (GLMM) function.

The necessary columns from the resultant file were then combined into a new file. This greatly reduced the size of the file. The function 'mice' was used for imputation of the data. This eliminates 'NA' by calculating the average mean for null data. Next, the 'sample' function was used to randomly select data from the reduced columns. The purpose of sampling was to train R to output decision trees. In doing so, appropriate decisions can be made by applying 'ctree', 'train' etc functions (a decision tree diagram is shown in appendix 1). The following 'range function':

$$\text{function}(x) \ \{(x - \min(x)) / (\max(x) - \min(x))\}$$

Was used to standardize the data such that they would lie between the range zero and one in the case of normal probability distribution. A natural logarithm was also used to make the data less skewed. A Gaussian GLMM and summary function used for observing the

results was important. There are three kinds of estimating distribution that can be used when attempting to fit the data – Poisson, Binomial and Gaussian. Based on the computing procedure and reasons described in the previous section, the data was standardized using the Gaussian method. The function 'Annova' was used for hypothesis testing concerning the relation between different variables. Finally, the function "prcomp" and "ggbiplot" was used to find the correlations between the different varying factors such as social objects (age, female, education, family incomes etc), cultural identity (knowing English, language spoken at home etc), and personality (cognitive) – the correlation diagram is shown in appendix 2 and 3. It is important to note that one can perform prediction through statistical simulation. This will be further discussed in the following reference [24].

Computing Results and their reasons

When talking about computational outcomes, one should first refer to Principal Component Analysis (PCA). This is used to understand the correlations between different variables (age, language, cognitive, female, language spoken at home etc) with reference to the categorical variables: do they have "access to Internet", do they own "a laptop" and do they "have a broadband connection"? These three factors were investigated because they were more likely to determine whether their relationships to social objects, cultural identity and personality.

First, in the case of having "access to the Internet" - (Appendix 2 and 3), it was discovered that most variables were correlated with each other. However, while some were positively correlated while others were negatively correlated. According to "Investopedia" [25], positive correlation means two variables are moving in tandem. This means when one variable is increasing in value, the corresponding variable is also increasing. The converse of decreasing is also true. Correlation shows that at least two variables are moving in the same directions. Thus, correlation is a form of dependency. In the field of psychology, positive psychological responses imply positive changes within an area. For the negatively correlated, a change in variable affects an opposite change in another variable.

This study found, (see appendix 2) that there were correlations between different variables with the factors of being able to access the internet, owning a broadband connection and having a laptop. Some are positively correlated variables, such as with the first group: eng, access, broadband and state. The reason being that most language used on the Internet likely to be English. It follows that when students learn more English, they can understand web content more easily and with lesser barriers. Therefore, they always own a broadband connection. Furthermore, there are also geographic factors to consider, for example in the

United states. The positive correlation for United States and Internet access might be because there is different state funding provided to rural areas [26]. Age, family's income, owning a laptop and household language are intra-positively correlated with each other but only slightly positive related with the first group. The reasons for this might be that young and non-white students are more likely to access the Internet than white students. Perhaps non-white families do not have access to other daily educational activities and as a result their children might try to use the Internet as a form of 'free entertainment'. In addition, non-white families might prefer to maintain friendships using Internet chat websites [27]. Hence, non-white young students with high family income prefer to own laptops. At the same time, there are also negatively correlated variables such as cognition, parent educational level, and race with the first group. With regards to cognition, the reason may be because not all internet websites allow reflection, analysis and imagination [28]. Most visual media is in real-time such as television and video game which do not have the aforementioned properties. Reading implies the development of imagination, induction, reflection and critical thinking among students. However, most young children like to use the Internet for chatting and gaming but not reading [29]. Therefore, the more they access Internet, the less cognitive for the students. In the case of the parent education, the lower their education, the more they will access Internet. The reason being they assume using Internet will result in a better education for their children. As a result, they prefer their children to have more access to the Internet. Race is also another negatively correlated factor because white people in United States pioneered the use of computers, with non-whites being the late comers [30]. Gradually however, non-white people grew to dominate access to the internet and richer white peoples shifted to other newly developed technology.

In brief of this study highlights why factors (age, income, state, parents' education, race etc), cultural identity (English, household language used etc.), and personality (such as cognitive etc.), are correlated using PCA methodology. Indeed, when students are more fluent in English, they can understand the web content more easily. Thus, if non-whites can acquire more English knowledge, this can help them to participate more in web educational activities rather than entertainment such as chatting. Furthermore, it is necessary to increase the proportion of reflection, analysis and imagination in different websites for students as well as more reading to increase cognitive among them.

DISCUSSION
Diversity in reading habits stems mainly from a country's culture and influences its educational philosophy. For example, in the Netherland, students are not taught to read until primary school. Their philosophy believe children should be in "a relatively, stress- free educational environment that emphasizes learning" [31]. The reason for such a philosophy comes from Dutch university studies. They explain their academic pressure is much lower compared with the rest of the world. In fact, both Dutch teachers and parents understand the significance of playing after school instead of only studying. The school learning environment is more relaxed and the quantity of homework is less compared with the amount received by a primary school student in the United States.

This results in a reduction of anxiety stress among students in the Netherlands, unlike in the United States and most Asian countries such as China, Singapore, South Korea. Although there are official tests by the government to primary school students, their selecting performance in secondary school will be respected. This means a child has the final say in choosing which type of school they want to attend to continue their studies. One of the most important factors is that there are enough university places which creates a less competitive environment.

Students are thus not pressured into achieving the best academic performance in class. Instead, it allows them to have more time to participate in extra-curricular activities such as dancing and hockey.

How can one solve the digital inequality problem? The following section gives a brief overview of social filter, and social control theory and explain why they are important.

Social filtering theory [32] concerns mating and dating selection. It tells us social structure is really the criteria for the number of partners when there is a mate. Hence, people like to date and marry those like them according to age, race, social status and religion. This is defined what we known as homogamous while for those who marry spouses with different characteristics are known as heterogamous. In other words, the spouse has complementing but not similar patterns.

Consider the example of "hip-hop" music. According to Perullo [33], artists like to compose music using their own local, social, and linguistic practices, then filter American and other foreign hip-hop styles. The result is they create their own distinct musical and artistic forms. In 1996, Bambaataa suggested that music is color blind and can transcend races. This leads to the problem of whether globalization will provide us a homogenized global culture or whether local culture will encourage heterogeneity [34]. Similarly, in this study, the key was to filter the homogenized behavior among students associated with social objects, cultural identity and personality prerequisite. By doing so, the misuse of information technology can be prevented, playing non-educational games for example, and the

digital equality in children can therefore be minimized. In other words, the reason to favor heterogamous behavior is mainly for its complementary aspects, such as using more ICT when learning.

It is important to note that, in mathematical simplex method, one may also have the elimination theory. Its aim is to eliminate and reduce the number of unknowns in simultaneous equations through multiplication, division, addition and subtraction with some numerical variables calculation process. Thus, one can use a similar approach by providing training to fulfill students' inadequacies in ICT usage. Then the homogenized behavior can be eliminated. With reference to the above example, by teaching students how to determine the usefulness of educational information in the web can minimize the digital equality in children. This also means, elimination can help students use more ICT in studying with a better approach and thus maximize the gain of digital usage outside schools.

Social controlling theory [35] concerns four types of social bonds linking parents to school society:

Attachment determines what level a human being is emotionally closed and sensitive to the other people [36]. For a low SES parent, this may include children, teachers and peers. Furthermore, the significance of emotional relationships with other peers needs to be considered. Indeed, low SES parents usually have similar expectations to their adolescents. Roughly speaking, with modification to Hirschi, 1969 [35], this study suggests using:

a) Supervision - the amount of time that parents spend with their children at home,
b) Communication - the flow of communication within the parent-teacher relationship and
c) Affectional identification - the degree of affection that exists between parents' peers.

Finally, when a parent has a positive attitude with school, then most of their delinquent behavior in supervising children can be eliminated [36]. It is because more attachment implies parents can get more help from school society in the matter of handling children's ICT usage request.

Commitment means whether an individual is willing to invest in school society as well as the extent of their attempts is trying to achieve educational goals. This refers to rational calculation in the potential gains and losses of behaviors when implementing goals. If low SES parents commit to school society, then they will find it easier to handle their children's ICT usage requests. This will be because they care more about investment in their children's expectation from society. Furthermore, commitment also refers to whether one is willing to conform to a society's standards. For example, when low SES parents invest less time and become less integrated in society, the monitoring of

their children also decreases since they feel the gains out-weighs the losses [35]. The reason being they may think financial problems are more serious than their children's education. Thus, education commitment is low and monitoring decreases.

Involvement refers to whether parents choose to actively participate in prosocial movements. This relates to the permission of children's ICT usage and how people decide to spend their time. Parents will know how to monitor their children when they spend time, energy and personal resources on school activities [37]. Indeed, the quality of these social activities and how many resources a person will spend determines their level of involvement. For instance, if one participates in a school training course concerning child' ICT usage, then they will have more of an idea in how to handle daily requests.

The final type of social bond is belief. This measures how much an individual accepts their society's moral norms, rules and laws. If one considers these values important, the chance that they will engage in delinquent behavior becomes decreases. In addition, the complexity of the developmental process should also be considered as this will impact belief.

Finally, the level of an individual's religiosity is another element of the belief social bond [38, 39]. More specifically, an example of belief is whether parents accept the general rules of ethical child ICT usage.

Other than social control theory, there is also self-control theory and rational choice theory. According to Hirschi and Gottfredson [40], one can refer to self-control theory as "crime theory". It is proposed that everyone will commit a crime but suffixed to that fact is there are also people who might not be vulnerable to momentary temptations. This theory can be applied in Lam, October, 2016 which states that parents' resistance to anger temptation can prevent family conflicts when using ICT. In addition, the rational choice theory states that an individual will prefer something that they think the best under the circumstances of potential rewards and costs. Indeed, most low SES parents value ICT usage but do not understand the significance of effective usage.

Therefore, it is the school's and society's responsibility to teach its importance. It is they who can make decisions about what are the best criterions for their child's ICT usage.

In addition to home ICT usage, there are also barriers in school teaching and learning of the educational technology and its uses. According to Hew and Bush in 2007 [41], there are several barriers in the pedagogy when one tries to integrate ICT into the curriculum. These are resources, subject culture,

institution, beliefs, and attitudes. One of the reasons for these barriers is because of the inaccurate estimation of the actual needs of ICT policy or its tools [22]. This may relate to the school's leadership, for example – the principal's vision.

According to Tan 2011 [42] there are three types of school leadership: top-down, segmentation and functional differentiation distribution. A school head may not need to know much about ICT but he should have the vision to integrate ICT into the curriculum. In addition, they could adopt a functional differentiation distribution model. The reasons for adopting such a model are:

1. The approach provides an in-depth level of distributed leadership which can reduce complexity [43]. It is thus easier for complex changes to occur in a school-wide ICT integration process.
2. Sub-units become more focused in their function, as well as the coordination between them [42]. This creates a stronger inter-dependency among the sub-units.
3. It becomes difficult for the leadership to suggest a path and supervise each sub-unit due to the complexity of the process. This results a higher level of empowerment at each level.

Firstly, the system is sub-divided into unequal sub-systems, each having a different function respectively. These sub-units rely on each other and form an independent network with essential communication between them. The Principal is willing to hand most of the ICT related work to the ICT department. The ICT head as well as the ICT coordinator work together to establish direction for ICT and technology related integration. The department consists of four sub-divisions, these are: student development, staff development, special projects and infrastructure. While the heads of the other department take care of various subject disciplines, they also work on the individual plans for departments' ICT integration. Other expert teachers are considered senior teachers and participate in planning key ICT programmes.

For example, in a specific English Department, students were asked to create story using digital media. This is a key pedagogy for language learning and acquiring skills in media literacy. The creation of the digital story required the students to use suitable software applications. The story also needed to contain text, sound recordings and digital images. Scaffolding tasks were given to pupils, where group brainstorm ideas and characters in the form of pairs for profiles, the outline of the story and their narration was recorded. Teachers' feedback was provided and all the digital stories were uploaded to the school network. The use of colorful visuals and music helped to excite and engage students. ICT can therefore facilitate in the presenting of digital stories which can be easily created and refined. At the same time, students can learn from one another during creation. While multiple readings and recordings were required, students were assisted in learning with ICT – through collaboration and production. For such a new policy, there was resistances from some of more experienced teachers. For example, when teaching ICT, they were against the use of Scratch as an introduction to programming and stated its disadvantages. After in depth discussion, they agreed to teach (test) it in one class and compare it with traditional programming such as Java. In a sense, they worked cooperatively as a whole for the test. In addition, there were also links crossing traditional boundaries as teachers from various subjects guided learning based on an integrated curriculum. For example, English provided lots of chances to use online discussion forums. Moreover, students were welcomed to participate in making school decisions.

For the successful integration of ICT to be implemented, teachers must first have an initiative sense and enough support. They should be encouraged to search for and find new kinds of teaching pedagogy. The Principal can play a much more critical role in the commitment, vision and belief in the use of ICT throughout the school. Besides promoting ICT curriculum integration, the Principal should commit more (fight for more ICT funding and provide suitable ICT pedagogy training for ordinary teachers) and share his belief (what direction should ICT pedagogy transfer to – using more online applications or introduce mobile devices) more with other staff. Finally, choosing a suitable person for ICT pedagogy leadership is key when effectively integrating support ICT in learning. In fact, an ICT coordinator, subject panels and senior teachers should work together in different aspects of leadership. From the above example, if a school can employ a distributed functional differentiation leadership, they will be successful in integrating ICT into the curriculum. Furthermore, to enhance successful implementation and brainstorming of ICT policy, leaders should establish effective communication with ordinary teachers' curriculum integration. Furthermore, to enhance successful of implementation and brainstorm of ICT policy, the leaders should have effective communication with ordinary teachers. In other words, communication acts as a bridge between leadership and the excellence of execution in ICT education. This can be achieved through collaboration [44] particularly online communication such as chats, discussion forum, and emails. The barriers caused by the school's leadership can be minimized together with the successful use of education technology. With better implementation of ICT policy, student academic results will no doubt be raised.

CONCLUSION
The statistics shows that there is a relation between social objects, cultural identity and personality. Social objects such as gender, affects reading habits in

different countries as well as the language gap between males and females. Furthermore, cultural identity such as reading habits have a connection with one's personality for example their learning behavior. All three factors will affect a student's digital usage or equality. In addition, there are also parental influences on ICT usage and equality. It has been shown that child ICT usage and equality effect their scholar results. Indeed, by supervising (controlling) the parental influence factor in terms of providing advices, the digital usage / equality of children can be affected. In doing so, the basic ideas about parents' attitude to handling the use of digital tools can be revolutionized. For example, philosophical education can be provided to parents allowing them to process their children's request for ICT usage. Next, the unfavorable factors among social objects, cultural identities and personality can be eliminated. This will have an influence on the digital usage / equality between students. For instance, one can eliminate cultural reasons for reading habits by providing digital games which encourage reading. Reading Battle developed by the University of Hong Kong Centre of Information Technology Education, can certainly provide motivation to Dutch students to read different kind of books.

Furthermore, the theories – social filtering theory, social control theory, self-control theory and rational theory can all be combined and applied to students and parents. They can be used to encourage positive and quality ICT usage after school, as well as maximize the gain from effective adolescent ICT usage at home. For examples, more school activities can be held for low SES parents to promote positive and quality ICT usage such that they have ideas in handling the request of children's ICT usage. In addition, more ICT educational digital products should be developed to help change students' daily ICT usage habits. Furthermore, when there is a good leadership and vision concerning ICT policy from a school's principal, the barriers which occurred from using technology in school teaching and learning can be minimized. For example, the government should encourage further study in ICT leadership and related knowledge among teachers and school authorities. Hence good leadership in schools implies the successful implementation of ICT education. This reduces the resistance of "positive and quality ICT usage" in teaching and learning between teacher and students. It can be concluded that optimization between student home ICT usage and school educational technology used in teaching and learning can result in the improvement of scholar achievement. In such a case, digital equality in education and students' academic performance will be sublimated into an optimization problem among:

1. How parents will handle children's ICT usage requests,
2. The effects of social objects, cultural identities and personalities to children's ICT usage, and
3. Schools' leadership in the policy of ICT usage.

Figure-4 in the next page will show us the framework of the above suggested systematic thinking.

Finally, how can one determine the optimal value in the situation for students, parents and schools in ICT usage? According to Harris, 2015, one can find out the value for each student through assessment tests. Indeed, these tests are used to reflect his / her learning outcomes. Then professionals and researchers can tailor made a schedule for each student such that he / she can best use the ICT products, prevent misuse and hence improve the academic achievement. Then the problem of digital equality in education will be overcome. At the same time, another new way to solve the equality problem is through education and learning [45]. This is because we are in the 21st Century

New Economies People are transformed with new innovations in their work under the effects of information technology. This means ICT is the main driver of Economic growth. Therefore, one needs the competence to exploit the machines and sophisticated digital devices. Hence, there is a need of lifelong education as well as learning such competence in schools and colleges.

In general, the proposed method is intensive and depends much on professionals or researchers' tailor schedule to individual student. There are more restrictions for students to follow. The new method is more extensive with less limitations since less rules should go with. Thus, the best suggestion is one should choose the solving method according to their own needs.

REMARKS
This author notes that the procedure that he used for handling those statistical big data in this paper is a kind of "tailor-made under a step by step" process. It can be further extended for the earthquake prediction that will be mentioned in the following section. This is known as a form of regression algorithm. They fall under the family of Supervised Machine Learning Algorithms.

Obviously, it is a subset of machine learning algorithms. One of the major applications of such an algorithm is using in modeling those dependencies together with the relationship between the target output and input features. Hence, one can predict the values for new data.

The most significant value of this paper is its computing procedure (in processing big data) which can be further evolved for an uncalibrated earthquake prediction. According to the scholar paper written by Asencio-Cortes, Morales-Esteban & Martinez-Alvarez [46], "generalized linear models (GLM) were finally used in combination with a model fitting based on the

AIC measure for predicting the probability of near-fault earthquake ground motion pules."

During the prediction, there was a set of four machine learning-based regressors:
1. Generalized Linear Model (GLM);
2. Gradient Boosting Machines (GBM);
3. Deep Learning (DL);
4. Random Forests (RF).

They were then combined with ensemble learning under the context of big data to achieve the ultimate purpose of prediction. In my paper, Principal Component Analysis, Generalized Linear Model, and Decision Tree Diagram were applied for the computation of results and so do the explanations. Intuitively, only the "Deep Learning" regressor (Deep Neural Network) was missed. In general, only GLM is the focus of this author's paper and is the same regressor with Gaussian distribution used in the Cortes *et al.*, [47] paper. It is no doubt that this author who develops an appropriate regression algorithm (RA) for the handling those big data during the research.

Certainly, it would be better if this author had used all the other two regressors together with DL for the final outcomes. However, my work had inspired the usage of (different) RAs in the prediction of earth-

quake. Besides, one may also use a database of past ground motions that either contains velocity pulse or its negation to calibrate the regression models for the prediction of such future pulses in ground motions [48]. Then the final outcomes will be more promising.

My generalized and extended regression algorithm processing procedure is:
1. Selection of data according to specific criteria,
2. Convert the file to numerical format from text if exist,
3. Combine columns into a new file,
4. Eliminate any "NA" by using data's mean value,
5. Random sampling (more than one time) from the immediate data,
6. Train the software for more than one output decision trees (GBM),
7. Standardization for Gaussian model,
8. Take logarithm for skewed,
9. Use Annova for the hypothesis testing,
10. Find those correlations,
11. Use wide sets of trees to perform further data predictions (RF),
12. Apply more sets of data for input and output (DL).
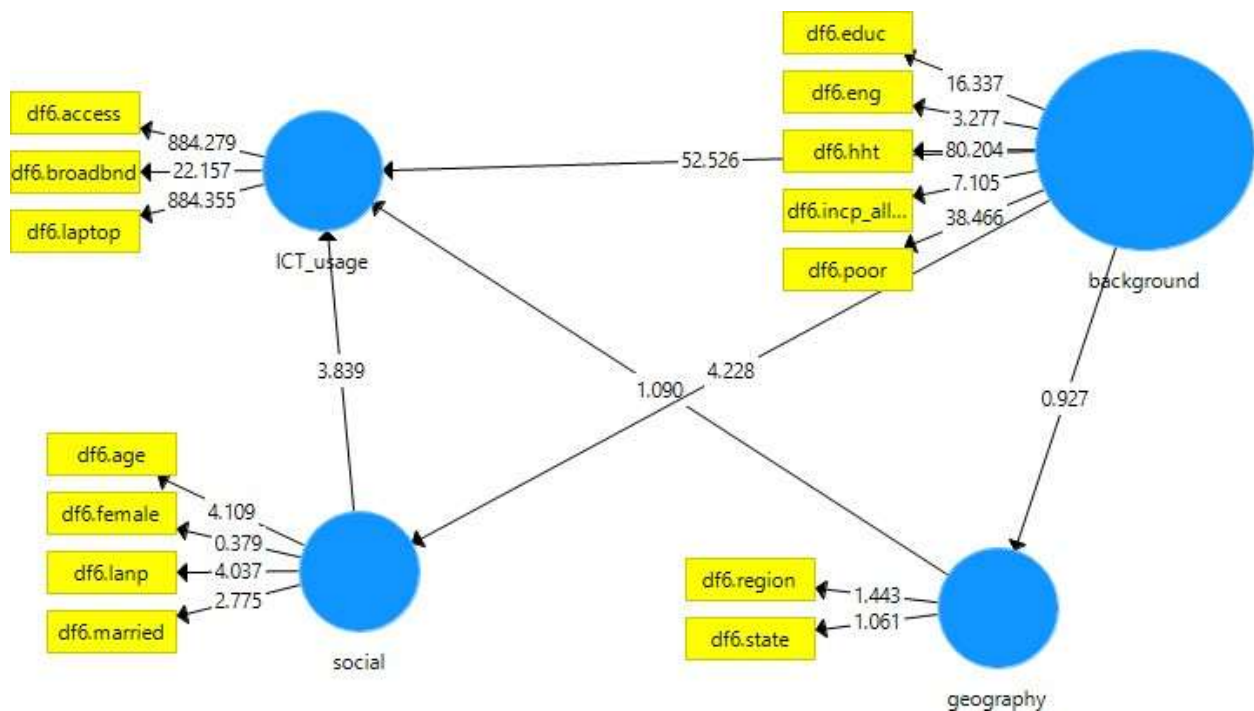13. Employ database to calibrate the Regression Model.

Fig-3

This author also remarks that the above Figure-3 is the Partial Least Square Structural Equation Model for the research. It should be note that according to the p values of specific indirect effects, social objects are acting as the mediator between the background and the ICT usage. This means a student's background and

social objects together with the ICT usage form a causal relationship. Or the social backgrounds do have effects in the ICT usage. The reason may be for those students in well educated family, they tend to use ICT for educational purpose under suitable guidance. While in the case of less educated family students, they tend to

use ICT for entertainment matters. This is because most of these low educated parents are without sufficient ICT knowledge in advising their children to use ICT in a non-entertainment way. Therefore, one should educate less privilege parents so that they may acquire suitable ICT knowledge in supervise and monitor their children's ICT usage.

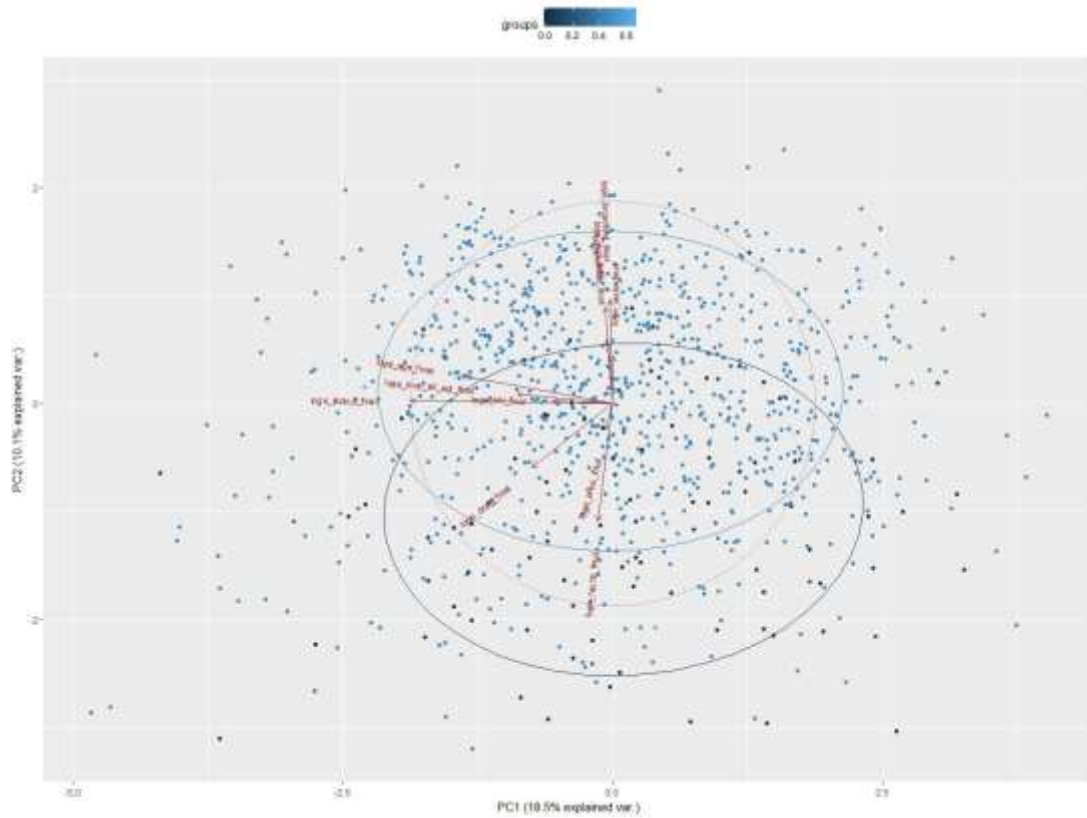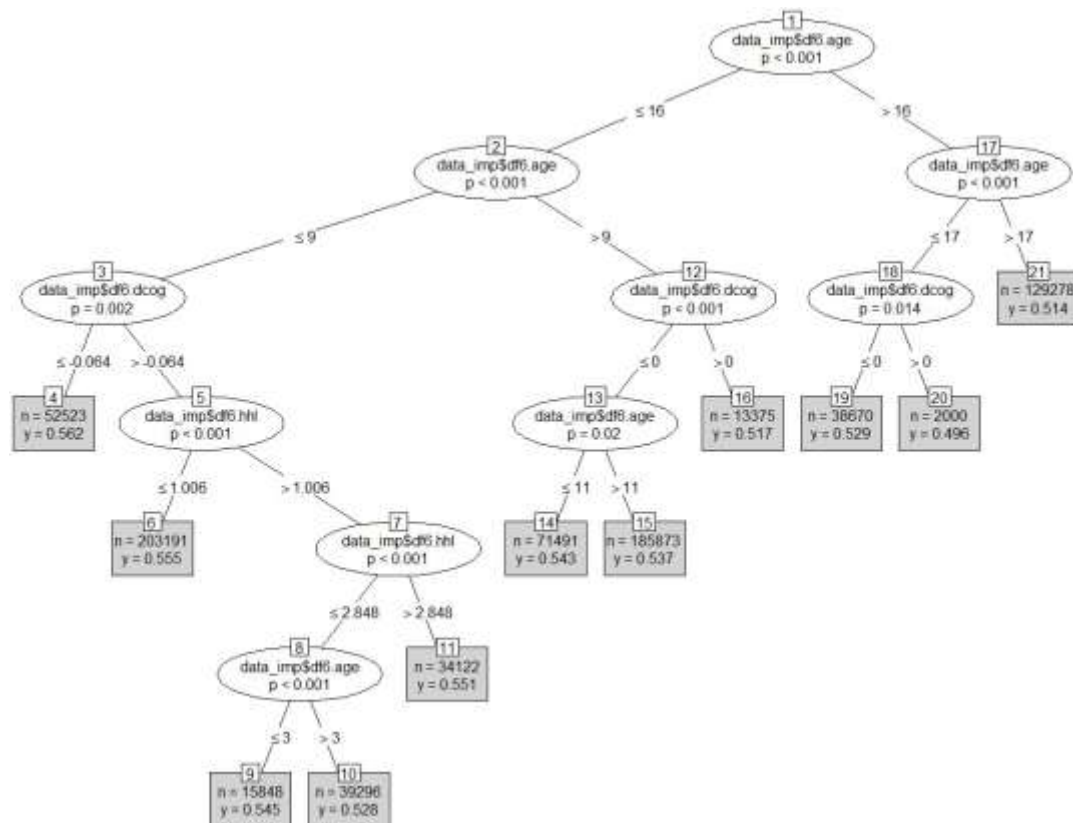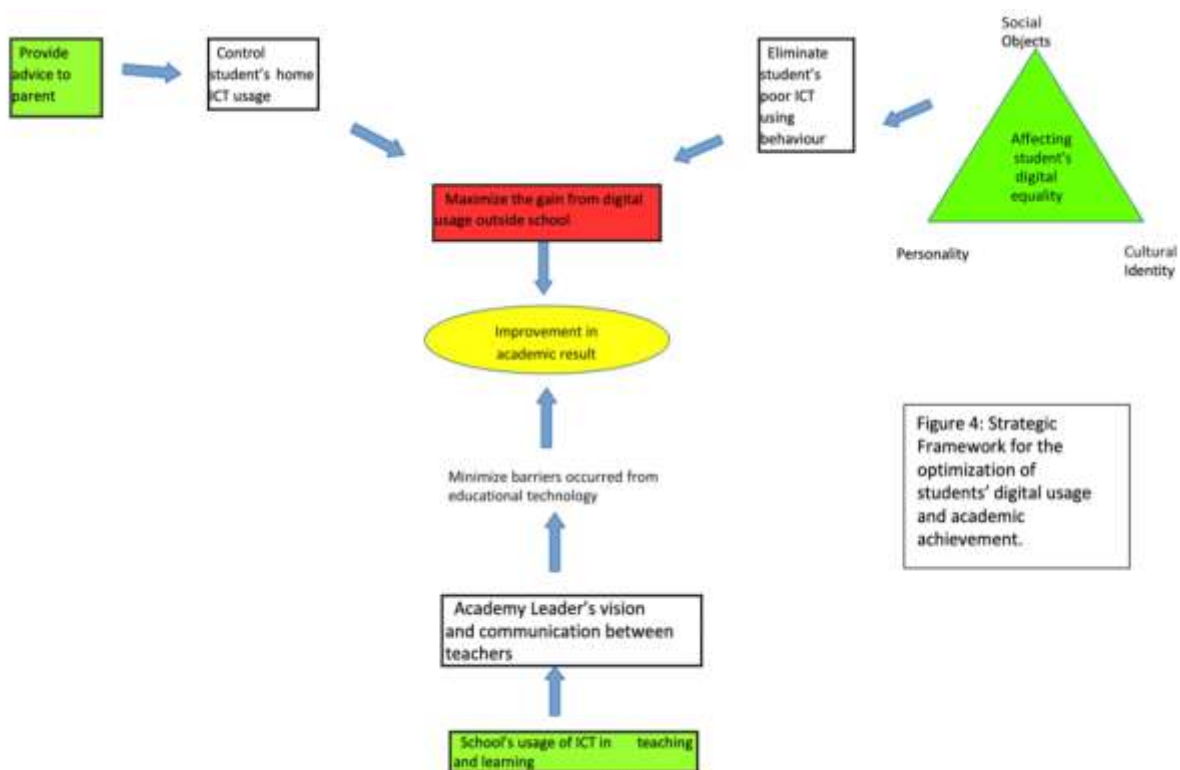Background —> Social Objects —> ICT Usage P value equals to zero

Fig-4

Appendix 2: The Correlation Diagram among Social Objects, Cultural Identity and Cognition in students' Internet access, having a laptop and owning a broadband connection.

Figure 4: Strategic Framework for the optimization of students' digital usage and academic achievement.

Appendix 3: The Probability Tree Diagram for correlating age, language used at home and cognition such that researchers can use it as a reference to make necessary decision in students' ICT usage

REFERENCES

1. Lam, S. K. (2016). The Need of Philosophy Education to Parents (May 3, 2016). Available at

SSRN: https://ssrn.com/abstract=2774598 or http://dx.doi.org/10.2139/ssrn.2774598

2. Lam, S. K. (2016). Cultural Transformations in Parents' Handling of Child ICT Usage (October 21, 2016). Available at SSRN: https://ssrn.com/abstract=2856892

3. https://www.analyticsvidhya.com/blog/2016/03/practical-guide-principal-component-analysis-python/

4. http://ordination.okstate.edu/PCA.htm

5. Gauch Jr, H. G., Whittaker, R. H., & Singer, S. B. (1981). A comparative study of nonmetric ordinations. *The Journal of Ecology*, 135-152.

6. http://philosophy.stackexchange.com/questions/37452/reading-for-philosophy-of-statistics-statistical-inference

7. Jauregui, C., Eidam, T., Otto, H. J., Stutzki, F., Jansen, F., Limpert, J., & Tünnermann, A. (2012). Physical origin of mode instabilities in high-power fiber laser systems. *Optics express*, 20(12), 12912-12925.

8. Elgamal, T., Yabandeh, M., Aboulnaga, A., Mustafa, W., & Hefeeda, M. (2015, May). sPCA: Scalable principal component analysis for big data on distributed platforms. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (pp. 79-91).

9. Shlens, J. (2003). A tutorial on principal components analysis: Derivation, Discussion and Singular Value Decomposition, Princeton University Press.

10. Walter, W. S. (2013). Generalized Linear Mixed Models: Modern Concepts, Methods and Applications, CRC Press, Taylor and Francis Group.

11. Jolliffee, I. T. (2002). Principal Component Analysis, Second Edition, Springer.

12. Williams, S. C. R. (2010). Progressive Lateral ventricular enlargement in schizophrenia: A metaanalysis of longitudinal MRI studies, Schizophrenia Research, 120(1-3):54-62.

13. Kieseppa, I. A. (2001). Statistical model selection criteria and bayesianism, Proceedings of the Philosophy of Science Association. (3):S141-S152. https://philpapers.org/rec/KIESMS-2

14. Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue internationale de philosophie*, 20-40.

15. Crosskey, W. W., & Jeffrey, W. (1980). *Politics and the Constitution in the History of the United States* (Vol. 1). University of Chicago Press.

16. Nuvunga, J. (2015). Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genetics and Molecular Research.* 14(4):14262-14278.

17. Robinson, G. K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6(1):15-32.

18. Mäntysaari, E., & Van Vleck, L. D. (1989). Restricted maximum likelihood estimates of variance components from multitrait sire models with large number of fixed effects. *Journal of Animal Breeding and Genetics (Germany, FR).*

19. http://ase.tufts.edu/gsc/gradresources/guidetomixedmodelsinr/mixed%20model%20guide.html

20. http://lme4.r-forge.r-project.org/book

21. http://bstt513.class.uic.edu/GLMM.pdf

22. Lau, K. G. (2014). Digital divide in education: a shift to ethical usage. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR. Retrieved from http://dx.doi.org/10.5353/ th_b5328043

23. Yuen, A. H. K., Lau, W. F., Park, J, Lau, K. K., & Chan, K. M. (2014). Home Computing and Digital Equity in Education: A Hong Kong Story, American Educational Research Association (AERA) 2014 Annual meeting. 3-7 April 2014, Philadelphia, Pennsylvania.

24. https://www.r-bloggers.com/probability-and-monte-carlo-methods/https://stats.stackexchange.com/questions/17730/writing-a-monte-carlo-simulation-in-r

25. http://www.investopedia.com/terms/p/positive-correlation.asp

26. http://www.rurdev.usda.gov/utp_bip.html

27. http://www.pewglobal.org/2015/03/19/internet-seen-as-positive-influence-on-education-but-negative-influence-onmorality-in-emerging-and-developing-nations/

28. http://newsroom.ucla.edu/releases/is-technology-producing-a-decline-79127

29. Lam, J. (2014). *Enterprise risk management: from incentives to controls*. John Wiley & Sons.

30. Jones, S., & Fox, S. (2009). *Generations online in 2009* (pp. 1-9). Washington, DC: Pew Internet & American Life Project.

31. http://www.findingdutchland.com/five-impressions-of-the-dutch-educational-system/

32. Regan, C. P. (2008). The mating game: a primer on love, sex, and marriage. SAGE. 40–41. Retrieved 12 October 2011.

33. Perullo A. Hooligans and heroes: youth identity and hip-hop in Dar es Salaam, Tanzania. *Afr Today*. 2005;51(4):75–101.

34. Callier, H. (2016). Hip-hop-ization: localizing a global youth culture: the remix. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR. Retrieved from http://dx.doi.org/10.5353/th_b5736674

35. Hirschi, T. (1969). Causes of Delinquency, University of California Press.

36. Weng, X. (2016). Social bonds, self-control, and perceived costs: an empirical investigation on Chinese juvenile delinquency. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR. http://hdl.handle.net/10722/235897

37. Lilly, J. R., Cullen, F. T., & Ball, R. A. (2011). The complexity of control: Hirschi's two theories and beyond Criminological theory: Context and consequences (5ed.). Thousand Oaks, Calif: Sage Publications.

38. Chapple, C. L., McQuillan, J. A., & Berdahl, T. A. (2005). Gender, social bonds, and delinquency: a comparison of boys' and girls' models. *Social Science Research*, 34(2):357-383.

39. Longshore, D., Chang, E., Hsieh, S. C., & Messina, N. (2004). Self-control and social bonds: A combined control perspective on deviance. *Crime & Delinquency*, 50(4):542-564.

40. Hirschi, T., & Gottfredson, M. R. (1993). Commentary: Testing the general theory of crime. *Journal of Research in Crime and Delinquency*, 30(1):47-54.

41. Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Educational technology research and development*, 55(3), 223-252.

42. Tan, S. C. (2011). Technology leadership and – Lessons from empirical research. In Steel, C., Keppell, M. J., & Gerbic, P. Curriculum, technology & transformation for an unknown future. Proceeding ascilite Sydney 2010.

43. Luhmann, N. (1995). Social systems: Standford University Press.

44. Carnell, E., & Lodge, C. (2004). Supporting Effective Learning. Paul Chapman Publishing.

45. https://www.oecd.org/site/schoolingfortomorrowknowledgebase/themes/ict/bridgingthedigitaldivide.htm

46. Asencio-Cortés, G., Martínez-Álvarez, F., Troncoso, A., & Morales-Esteban, A. (2017). Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks. *Neural Computing and Applications*, 28(5), 1043-1055.

47. Cortes, H. (1908). *Letters of Cortes* (Vol. 1). GP Putnam's sons.

48. Shahi, S. K., & Baker, J. W. (2011). An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, 101(2), 742-755.