Ade Dedi Rohayana, Umi Mahmudah
Indonesia

The New Educational Review

Canonical Correlation for Analyzing the Relationship between Educational Attainment and Marriage of Indonesian Youths

Abstract

Youth plays a significant role for the future of a country in facing the rapid changes of time. This study examines the relationship between educational attainment and age of the first marriage of Indonesian youth. Data concerning Indonesian youth published by Statistics Indonesia in 2015 is used, which includes as many as 34 provinces of Indonesia. This study uses five age groups of first marriage as dependent variables (y). There are also five groups of educational attainment that are used as independent variables (x). Canonical correlation analysis is applied to analyze the strength of correlation between these two data matrices by using R program. Results provide five functions, which reveal that canonical correlations of these variables are 0.82, 0.60, 0.32, 0.16, and 0.01. Furthermore, the first and the second functions accommodate as much as 74.45% and 20.36% of the canonical relationship, respectively. Meanwhile, the rest is accommodated by the other three functions. The results also reveal that early marriage of Indonesian youth has a negative relationship with educational attainment.

Keywords: *educational attainment, age of first marriage, youth*

Introduction

Indonesia as a developing country puts high expectation on youth in order to compete globally. According to the Law of the Republic of Indonesia Number 40 Year 2009 on Youth, youth includes Indonesian citizens aged 16–30. All young Indonesians, whether they are students or have completed their education, are the foundation of the nation in actualizing the ideals of the nation as well as maintaining Indonesian supremacy in the world. Other than that, young citizens are expected to be agents of change in order to improve the nation's competitiveness in the international arena. Therefore, Indonesian youth character must be strengthened and empowered to provide strong leadership character so that they can carry out their duties properly. It is well known that youth faces a complexity of problems in their lives, which are a huge challenge for the country in providing the right direction so that their abilities can be maximized for the betterment of the country. Quality education is a good way to give youth a chance to prepare themselves for future competition. Therefore, educational attainment of youth can be used as reference of their quality. Meanwhile, Indonesia also faces a serious problem regarding marriage at an early age, where based on the national socio-economic survey from 2008–2012, 17% of women who got married at the age between 20–24 reports that they got married before the age of 18 years (Statistics Indonesia, 2015). The following figure shows the fluctuation of the percentage of married women aged 20–24 years where they got married before 19 years.
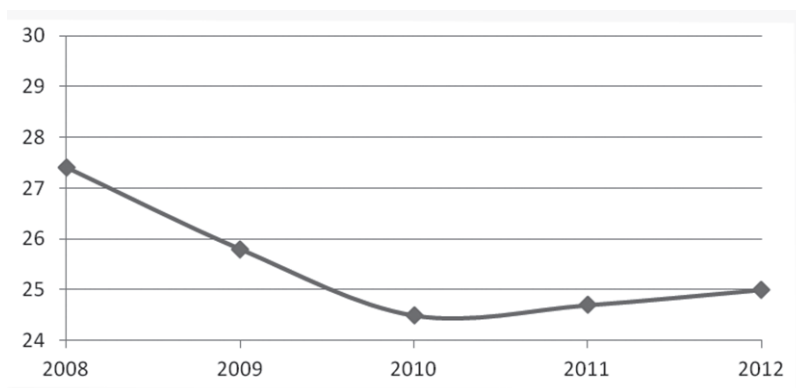

Figure 1. Percentage of married women aged 20–24 years who got married before the age of 19
Source: Statistics Indonesia, 2015

Statistics Indonesia (2015) reports that based on the national socio-economic survey in 2015, the population of Indonesia was approximately 254.9 million people, where as many as 61.68 million people (24.20%) was youth. Further, the percentage of youth by age groups can be explained as follows. As much as 32.38% (or around 19,969.28) of youth are aged between 16–20 years, whereas those between 21–25 and between 26–30 years of age amount to 33.51% (or 20,672.42) and 34.11% (or 21,041.61), respectively. Meanwhile, the majority of youth (55.79%) is not married, as compared to married youth, at 42.64%, whereas the rest are those who are divorced or dead. Figure 2 shows the percentages of married youth based on the age of first marriage and gender in 2015.
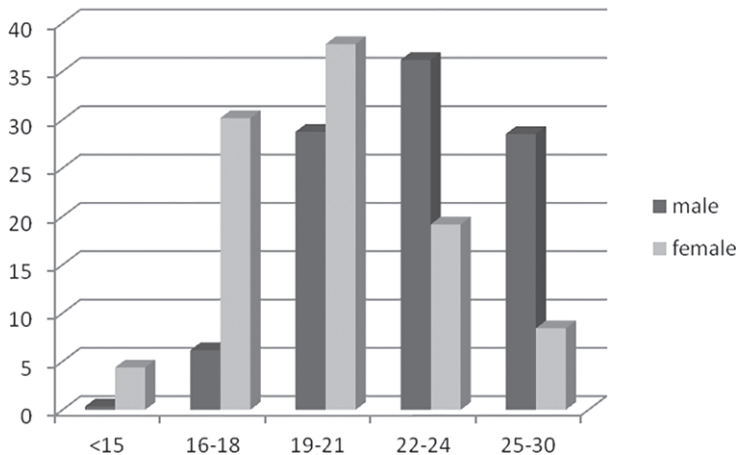

Figure 2. Percentages of age of first marriage and gender
Source: Statistics Indonesia, 2015

Figure 2 indicates that young women tend to get married at an earlier age, compared to young men. Further, early marriage is also dominated by female youth, where the percentages of married girls and married boys who married for the first time at the age under 15 years are 4.39 and 0.36, respectively. Meanwhile, the percentage of young women who got married for the first time at the age group of 16–18 years is about 30.18, compared to 6.16 of male youth. It is important to note that early marriage in Indonesia is regulated in marriage laws, where article 7 states that the permitted age of the bride and the groom are 16 years and 19 years, respectively. Figure 2 also shows that the majority of male youth (36.19%) tends

to get married for the first time at the age between 22–24 years. Table 1 presents the percentage of Indonesian youth based on gender and educational attainment in 2015.

Table 1. Percentage of Indonesian youth

Education	Male	Female	Male + Female
Never go to school	1.04	1.25	1.14
Do not pass primary school	5.35	3.82	4.60
Primary school	20.22	19.47	19.85
Junior high school	32.78	33.73	33.25
Senior high school and upwards	40.61	41.73	41.16
Total	100.00	100.00	100.00

From Table 1 it can be seen that the percentage of Indonesian youth who has senior high school diplomas and upwards is about 41.16, whereas 33.25% of Indonesian youth has graduated from junior high school. Further, there are 4.60 and 1.14% of Indonesian youth who has not completed primary school and never gone to school, respectively. Thus, there is 5.74% of youth who does not fulfill the obligation of the nine-year education. It is important to note that the Indonesian government implements a nine-year compulsory education program for all its citizens by issuing the Law of National Education Number 2 of 1989.

The main purpose of this study is to analyze the correlation between the educational attainment and age of first marriage of Indonesian youth. This study uses canonical correlation in order to analyze two sets of variables, where both the educational attainment and age of first marriage cover as many as five variables. Canonical correlation was first introduced by Hotelling (1936). Borga (2001) states canonical correlation is used to measure the relationship between two multidimensional variables, where one variable is treated as a dependent variable and the other is treated as an independent variable.

There are studies that investigate the relationship between education and marriage. Torr (2011) analyzes the relationship between education and marriage by using multinomial logistic regression (in the United States), which focuses on the changing relationship between education and marriage for both white and black women. Results reveal that education had a negative relationship with marriage for both white and black women in 1940, but they had a positive relationship in 2000. Musick, Brand and Davis (2012) also investigate the relationship between

education and marriage, using data from a national longitudinal survey of youth by applying propensity score analysis. Their results show that the effect of college attendance on marriage is significantly increased.

Meanwhile, canonical correlation is used in the context of student score analysis (Dai et al., 2011). Nanayakkara and Peiris (2017) use canonical correlation to analyze the effect of a mathematics course on engineering programs in Sri Lanka, which uses as many as 626 engineering students at the University of Moratuwa. Their results show that mathematics performance has a significant influence on engineering students.

Therefore, it is necessary to explore the relationship between the educational attainment and age of first marriage of youth in Indonesia. Thus, this study can be used as one of the references providing the right direction for Indonesian youth in order to improve their qualities by improving educational attainment, which is expected to have an inverse relationship with early marriage. Further, this study is expected to provide an overview of early marriage in Indonesia and its relationship with the level of education.

Research Methodology

In order to analyze the correlation between the educational attainment and age of first marriage of Indonesian youth, this study uses as many as 34 Indonesia's major political subdivisions, where the dependent variable consists of five groups of the age of first marriage, i.e., the age group under 15 years ($y1$), the age group of 16–18 years ($y2$), the age group of 19–21 years ($y3$), the age group of 22–24 years ($y4$) and the age group of 25–30 years ($y5$). There are also five variables in the set of independent variable, which represent educational attainments, i.e., youth who has never gone to school ($x1$), youth who has not graduated from primary school ($x2$), youth with a primary school diploma ($x3$), youth with a junior high school diploma ($x4$), youth with senior high school diplomas and upwards ($x5$).

This study uses canonical correlation to investigate the association between the educational attainment (x) and age of first marriage (y) of Indonesian youth, where both the dependent and independent variables consist of five variables. The main purpose of this analysis is to measure the correlation level of these two multidimensional variables and then describe the relationship structure in both the dependent and independent variables. Rencher (2002) states that this analysis is the development of multiple correlation, where there is one dependent

variable (y) and several independent variables (x). According to Hair et al. (2010), canonical correlation can be elaborated as follows. There is a group of independent variables (x_1, x_2, …, x_m) and there is a set of dependent variables (y_1, y_2, …, y_n). There is a new indicator from linear combination n, which is defined by P so that Q is the new indicator from linear combination m, where these combinations have maximum correlation, so that it can be written as follows:

$$P = a'x = a_1x_1 + a_2x_2 = \ldots a_mx_m$$
$$Q = b'y = b_1y_1 + b_2y_2 = \ldots b_ny_n$$

(1)

From equation (1) the following equations are obtained.

$$var(P) = a'Cov(x)a = a'\sum_{xx} a$$

$$var(Q) = b'Cov(y)b = b'\sum_{yy} b$$

(2)

$$Cov(P,Q) - a'Cov(x,y)b = a'\sum_{xy} b$$

Therefore, the correlation between P and Q can be written as follows:

$$Corr(P,Q) = \frac{Cov(P,Q)}{\sqrt{var(P)}\sqrt{var(Q)}}$$

$$Corr(P,Q) = \frac{a'\sum_{xy} b}{\sqrt{a'\sum_{xx} a}\sqrt{b'\sum_{yy} b}}$$

(3)

Furthermore, Hair et al. (2010) report that there are three methods in interpreting canonical correlation analysis, which are canonical weight, canonical loading and canonical cross loading. The first method describes the magnitude contribution of the original variable in its canonical variables, while the second method measures the correlation between dependent and independent variables and its canonical variables, whereas the third method illustrates the correlation between the original variables in one group and its canonical variables in another set.

Research Results

This study employs the data statistics of Indonesian youth of 2015, published by the Sub-Directorate of Social Welfare and Education Statistics of Statistics Indonesia, which are the results of the national socio-economic survey and the national survey of Social and Culture and Education Module in 2015. The publication covers the condition of Indonesian youth, which is conducted every year. A brief description of youth in Indonesia is presented, which includes the demographic structure, education, health, employment, access to media, and socio-cultural activities of Indonesian youth. However, this study focuses on the association between the educational attainments and age of first marriage of Indonesian youth by applying canonical correlation analysis. All of the 34 provinces are analyzed using R program in order to comprehend the relationship of the two sets of dependent and independent variables. Table 2 presents the descriptive statistics of the expected variables, whereas Tables 3, 4 and 5 describe correlations among the original dependent and independent variables.

Table 2. Descriptive statistics

Variables	Minimum	Maximum	Mean	Std Deviation
Dependent variables				
$y1$	0.50	6.25	3.23	1.42
$y2$	8.85	29.52	21.28	5.04
$y3$	27.64	38.40	34.08	2.40
$y4$	19.32	32.14	25.36	3.18
$y5$	9.67	30.95	16.05	4.57
Independent variables				
$x1$	0.41	24.24	1.76	4.05
$x2$	0.63	19.86	6.43	3.92
$x3$	7.84	28.99	18.85	5.41
$x4$	22.77	40.78	30.18	3.94
$x5$	26.82	66.30	42.77	9.77

Table 3. Correlations of age of first marriage

Variables	$y1$	$y2$	$y3$	$y4$	$y5$
y1	1.00	0.88	0.09	-0.86	-0.73
y2	0.88	1.00	0.33	-0.94	-0.90

Variables	$y1$	$y2$	$y3$	$y4$	$y5$
y3	0.09	0.33	1.00	-0.38	-0.65
y4	-0.86	-0.94	-0.38	1.00	0.80
y5	-0.73	-0.90	-0.65	0.80	1.00

Table 4. Correlations of educational attainment

Variables	$x1$	$x2$	$x3$	$x4$	$x5$
x1	1.00	0.14	0.05	-0.38	-0.35
x2	0.14	1.00	0.54	-0.31	-0.64
x3	0.05	0.54	1.00	0.25	-0.90
x4	-0.38	-0.31	0.25	1.00	-0.26
x5	-0.35	-0.64	-0.90	-0.26	1.00

Table 5. Correlations between educational attainment
and marriage

Variables	$y1$	$y2$	$y3$	$y4$	$y5$
x1	0.00	0.07	0.26	-0.07	-0.17
x2	0.51	0.54	0.36	-0.63	-0.50
x3	0.58	0.68	0.44	-0.70	-0.68
x4	-0.06	0.15	0.21	-0.05	-0.22
x5	-0.50	-0.68	-0.59	0.69	0.74

The results of correlations among the original variables show that the correlations between the educational attainment and age of first marriage are not excessive, where the largest value is 0.74 between the age group of 25–30 years ($y5$) and youth with senior high school diplomas and upwards ($x5$), as presented in Table 5. Therefore, from Table 3 it can be seen that the largest correlation of the within-set of dependent variable is 0.94 (negative correlation) between the age group of 16–18 years ($y2$) and the age group of 22–24 years ($y4$). Further, Table 4 indicates that the largest correlation of the within-set of independent variable is 0.90, which is a negative correlation between the youth with a primary school diploma ($x3$) and the youth with senior high school diplomas and upwards ($x5$), whereas Tables 6 and 7 illustrate the canonical coefficients of the dependent variable y and independent variable x.

Table 6. Canonical coefficients of dependent variables

Variables	Ycan1	Ycan2	Ycan3	Ycan4	Ycan5
$y1$	29.83	8.54	-15.27	18.94	-162.12
$y2$	29.98	7.12	-14.58	17.48	-162.17
$y3$	30.06	7.49	-14.34	18.24	-162.09
$y4$	29.74	7.10	-15.08	17.99	-162.18
$y5$	29.91	7.61	-14.30	17.77	-162.20

Table 7. Canonical coefficients of independent variables

Variables	Xcan1	Xcan2	Xcan3	Xcan4	Xcan5
$x1$	5.40	36.72	136.52	-50.29	74.31
$x2$	5.49	36.90	136.73	-50.46	74.02
$x3$	5.43	36.95	136.23	-50.49	74.38
$x4$	5.43	36.63	136.64	-50.57	74.15
$x5$	5.34	36.88	136.44	-50.47	74.23

Tables 6 and 7 show canonical coefficients which define the relationship between the variables in one group and its canonical variations. The dependent variable has five canonical variations, i.e., Ycan1, Ycan2, Ycan3, Ycan4, and Ycan5 and the independent variable also has five canonical variations, i.e., Xcan1, Xcan2, Xcan3, Xcan4, and Xcan5. It is important to note that these canonical coefficients have similar meaning to the regression coefficients by assuming that the outcome variable is the canonical variation. Meanwhile, Table 8 indicates the correlations between canonical variates of the dependent variable *y* and independent variable *x*.

Table 8. Canonical correlation analysis

Roots	Canonical Correlation	Squared Canonical Correlation	Eigen Values	Percentages	Pr(>F)
1	0.82	0.67	2.03	74.45	0.01
2	0.60	0.36	0.56	20.36	0.47
3	0.32	0.10	0.12	4.27	0.93
4	0.16	0.02	0.02	0.91	0.95
5	0.01	0.00	0.00	0.00	0.98

Table 8 shows canonical correlations which are correlations for the canonical variate pairs, where the first canonical correlation is 0.82, which is the largest value among the others, while the last canonical correlation is the smallest value (0.01). Further, the second pair has a correlation coefficient of 0.60, while the correlation coefficient of the third and fourth pairs are 0.32 and 0.16, respectively. In order to show the relationship between early marriage and educational attainment, Table 9 presents coefficient regression where the dependent variable is defined by the variable of the age group under 15 years ($y1$), while the independent variable are the five educational attainments. Meanwhile, Table 10 indicates the coefficient regression, where the dependent variable is the age group of 16–18 years ($y2$) by using the same independent variables.

Table 9. Coefficient regression of dependent variable $y1$

	B	Std. Error	t	Sig.
(Constant)	16.07	6.14	2.62	0.01
$x1$	-0.18	0.08	-2.18	0.04
$x2$	-0.07	0.12	-0.60	0.55
$x4$	-0.20	0.11	-1.88	0.07
$x5$	-0.14	0.05	-2.59	0.01

Table 10. Coefficient regression of dependent variable y2

	B	Std. Error	t	Sig.
(Constant)	47.93	19.74	2.43	0.02
$x1$	-0.36	0.27	-1.33	0.19
$x2$	-0.02	0.38	-0.06	0.95
$x4$	-0.24	0.35	-0.69	0.50
$x5$	-0.44	0.17	-2.54	0.02

Tables 9 and 10 show similar results, where all the independent variables provide negative values of coefficient regression, which indicates a negative relationship between the educational attainment and early marriage of Indonesian youth. However, there is one excluded variable, i.e., youth with a primary school diploma ($x3$), due to a collinearity problem.

Discussion

Analysis of the results shows that the correlation matrix of the dependent variable, which is defined by the ages of first marriage of Indonesian youth, is 5x5 because there are five variables in this group. The dependent variables are highly correlated, where the highest correlation is 0.94 (between the age group of 16–18 years and the age group of between 22–24 years). Furthermore, there is a high correlation between the age group of 16–18 years and the age group of 25–30 years, which is 0.90. High correlations are also indicated by the variables of the age group of under 15 years and the age group of 16–18 years (0.88), the age group of 22–24 years and the age group of 25–30 years (0.80). Meanwhile, the lowest correlation was found between the variables of the age group of 16–18 years and the age group of 19–21 years (0.33). The results of these correlations influence the summary of this group as a linear combination of dependent variables.

Correlations among the original independent variables provide different results with dependent variables, as can be seen in Table 4, where there is only one correlation coefficient close to 0.90, which is the association between youth who has diplomas of primary school and senior high school and upwards. The second highest correlation is 0.64 (between youth who has not graduated from primary school and youth with a diploma of senior high school and upwards). Further, the results also obtain the 5x5 correlation matrix as his group consists of five variables. Overall, the variable of the age of first marriage has a higher correlation than the educational attainment.

Furthermore, the analysis results in Table 6 reveal that the canonical coefficients for all the dependent variables have values that are not much different from one another regarding each variation of the variable of the age of first marriage. Based on Table 6, it can be explained that a one-unit increase in the age group under 15 years ($y1$) provides as many as 29.83 units of increase in Ycan1, which is the first variation of the age of first marriage (y). A one-unit increase in this variable ($y1$) leads to 8.54 unit increase in the second variation of dependent variable (y) and a 15.27 unit decrease in the third variate of y. The age group of 19–21 years has the largest canonical coefficient, i.e., 30.06, while the age group of 22–24 has the smallest value, i.e., 29.74, in the first variation of dependent variable y. Further, the third and fifth canonical variations provide negative values for all the elements in dependent variable y. Table 7 indicates canonical coefficients for independent variable x, where there is only one variation, i.e., the fourth variation (Xcan4) obtains negative values for all variables x. The third canonical variation (Xcan3) provides larger coefficient values than the other ones, where the youth who has

not finished primary school ($x2$) has the largest coefficient, i.e., 136.73. Further, a unit increase in youth who has not finished primary school ($x2$) leads to 5.49 and 36.90 units increase of in the first and the second canonical variations of the educational attainment, respectively.

From Table 8 it can be seen that there are five roots that indicate the first function to the fifth function, where canonical functions are chosen based on the eigenvalues from the largest to the smallest values. However, these values are calculated by using the values of squared canonical correlation. For example, the first root provides the eigenvalue of $0.67/(1-0.67)$, which equals 2.03, whereas the second root has the eigenvalue of $0.36/(1-0.36) = 0.56$, etc. The squared canonical correlation indicates that the variance proportion of canonical variation in one group is explained by the canonical variation of the other group. In other words, these values have similar meaning to R squared in regression analysis, where the first value of squared canonical correlation is calculated by $(0.82*0.82) = 0.67$. Further, the sizes of eigenvalues are associated with the correlation tests where the higher eigenvalues provide lower p-values, which is associated with F value. Therefore, the higher the canonical correlations have higher probability to be significantly different from zero.

Furthermore, the canonical correlations for the five roots are presented in Table 8, where the first and the second pairs have coefficient value of 0.82 and 0.60, respectively. Meanwhile, the third pair has the correlation coefficient of 0.32 and the coefficient values of the fourth and the fifth variations are 0.16 and 0.01, respectively. From Table 8 it can be seen that the first function accommodates as much as 74.45%t of canonical relationship, while the second function accommodates 20.36% of canonical relationship. There is 4.27 % of canonical relationship, which is accommodated by the third root, while the fourth and the fifth functions only accommodate as much as 0.91 and 0.01% of canonical relationship, respectively. Based on these results, the first function is more meaningful than the other functions. However, the second function also has a considerable contribution.

As mentioned before, early marriage for female and male youth in Indonesia is considered when the ages of the bride and the groom are 16 and 19 years, respectively. From Table 5 it can be seen that early marriage as represented by the ages under 15 years old ($y1$) has a zero correlation with the variable of $x1$ (youth who has never gone to school), while the correlation coefficients between this variable and the variable of youth who has not graduated from primary school ($x2$) and who has a primary school diploma are 0.51 and 0.58, respectively. Further, the correlation coefficient provides negative values between the variables and youth

with diplomas of junior and senior high school and upwards, which are -0.06 and -0.50, respectively.

Meanwhile, Tables 9 and 10 indicate that early marriage has a negative relationship with all the variables of educational attainment. However, this study uses two variables in order to define early marriage in Indonesia, which states that the bride's and the groom's ages have to reach 16 and 19 years, respectively. Therefore, these variables are the ages under 15 years ($y1$) and the age group of 16–18 years ($y2$). Table 9 shows the regression estimates, where the dependent variable is the ages under 15 years ($y1$), which indicates that the two variables have a significant negative effect on early marriage ($y1$), which concerns the youth who has never gone to school ($x1$) and those who have diplomas of senior high school and upwards ($x5$). Further, when the youth of the age group of 16–18 years ($y2$) is used as the dependent variable, then, based on Table 10, it can be seen that there is only one variable significantly influencing early marriage ($y2$), which is the youth with diplomas of senior high school and upwards ($x5$). Overall, it is safe to conclude that the early marriage of Indonesian youth has an inverse relationship with educational attainment. This means that by increasing the level of education of youth, early marriage rates can be reduced. This result is in line with previous studies, which report that education and marriage have an inverse relationship (Musick et al. (2012), and Torr (2011)).

Conclusions

This study explores the correlation between the educational attainment of Indonesian youth and the ages of their first marriage by using canonical correlation analysis, which is very useful in investigating the relationship between two dimensional variables. There are five variables in both dependent and independent variables, which are analysed using R program. The results reveal that there are five functions accommodating the canonical relationship but the first and the second functions have higher meaning, accommodating as much as 74.45% and 20.36% of canonical relationship, respectively. Furthermore, the results also reveal that all the variables in the educational attainment negatively influence early marriage of Indonesian youth.

References:

Borga, M. (2001). *Canonical Correlation: a Tutorial*. http://people.imt.liu.se/˜magnus/cca/, retrieved on 16 March 2018.

Dai, L., Chen, J., Li, S., and Dai, S. (2011). *Application of Canonical Correlation Analysis in Student Score Analysis Based on Data Analysis*. In: Lin S., Huang X. (eds.). Advances in Computer Science, Environment, Ecoinformatics, and Education. CSEE 2011. Communications in Computer and Information Science, 217. Springer, Berlin, Heidelberg

Hair, J., Anderson, R., Tatham, R., and Black, W. (2010). *Multivariate Data Analysis with Reading 7th Edition*. New Jersey: Prentice Hall International Edition.

Hotelling, H. (1936). Relations between Two Sets of Variates. *Biometrika*, 28, 321–377.

Musick, K., Brand, J.E., and Davis, D. (2012). Variation in the Relationship Between Education and Marriage: Marriage Market Mismatch?. *Journal of Marriage and Family,* 74(1) https://doi.org/10.1111/j.1741–3737.2011.00879.x

Nanayakkara, K.A.D.S.A. and Peiris, T.S.G. (2017). Identifying the Influence of Mathematics on Academic Performance of Engineering Students. *Engineering Research Conference (MERCon) 2017 Moratuwa*, pp. 247–252.

Rencher, C.A. (2002). *Methods of Multivariate Analysis 2nd Edition*. Canada: John Wiley and Sons.

Statistics Indonesia. (2015). *Statistik Pemuda Indonesia 2015* (*Indonesian Youth Statistics 2015*). https://microdata.bps.go.id/mikrodata/index.php/citations/407, retrieved on 16 March 2018.

Torr, B.M. (2011). The Changing Relationship between Education and Marriage in the United States, 1940–2000. *Journal of Family History*, 36(4), 483–503.