ICIC Express Letters
Part B: Applications
Volume **12**, Number **6**, June **2021**

pp. **507–514**

FUZZY HIERARCHY OF THE CORRELATION BETWEEN EDUCATIONAL INVESTMENT AND STUDENT ACADEMIC PERFORMANCE

Yi-Cheng Lin[1], Ya-Ling Lu[2], Ber-Lin Wu[3] and Sung-Po Lin[4,*]

[1]Kang Shou Elementary School
No. 269, Nanyang Road, Nantou City, Nantou County 54062, Taiwan

[2]Te Shing Elementary School
No. 200, Zhongxing Road, Nantou City, Nantou County 54041, Taiwan

[3]Department of Mathematical Sciences
National Cheng Chi University
No. 64, Sec. 2, Zhinan Road, Wenshan District, Taipei City 11605, Taiwan

[4]Department of Educational Policy and Administration
National Chi Nan University
No. 1, Daxue Road, Puli Township, Nantou County 54561, Taiwan
*Corresponding author: splin@ncnu.edu.tw

Received October 2020; accepted January 2021

Abstract. *Researchers have proposed numerous methods for detecting and testing correlations. In practice, the correlations within a data set may change with a change in regime. Regime change analysis can thus aid the fitting of different models to different regimes of data when conducting an economic interpretation of the data in a certain regime. This study presents an integrated identification procedure for change points/period detection. The membership function of each data point corresponding to cluster centers is calculated as a performance index grouping. Finally, an empirical study is conducted on change points/period identification for educational investment and performance.*
Keywords: Academic performance, Educational investments, Fuzzy correlation coefficient, Fuzzy hierarchy of correlation

1. **Introduction.** Academic performance is not only prized by parents and teachers but also an indicator of school quality. In the context of the emphasis on fostering a competitive workforce, academic performance and student ability are perceived as important indicators of social progress. According to the long-term research findings of Program for International Student Assessment (PISA), which is organized by the Organization for Economic Cooperation and Development (OECD), the relationship between socioeconomic status and educational achievement is significant, and countries also differ in their quality of education [1]. Children from low-income and low-educated families face more learning obstacles due to a relative lack of educational resources [2,3].

Factors affecting student performance have become an issue of concern for scholars in various fields. Scholars in Taiwan and abroad have differed in their findings on the determinants of academic performance, including teachers' politeness [4], intelligence, family environment [5], participation of parents in their children's education [6], and family socioeconomic background in the domains of social capital, cultural capital, and economic capital [7,8]. However, parents with the same capital may not invest the same level of resources into their children's education, indirectly leading to differences in academic performance [9].

DOI: 10.24507/icicelb.12.06.507

The degree of parental investment in education in most families is perceived not as a definite number but as a fuzzy one, which can be modeled as a continuous fuzzy interval. Human thinking is necessarily fuzzy because of fluctuations in our subjective awareness of time, status, and the environment [10]. Fuzzy theory is based on fuzzy logic, which extends the bivalence of classical logic (underpinning most mathematics) and upends the binaries of right/wrong and true/false. Yang et al. used fuzzy AHP to establish the hierarchical structural model of driver route choice, and the results were close to the real-world travelers [11]. In this study, we propose an approach to detecting trends and periods of change through fuzzy statistics by employing the K-means method. We identified the disturbance and discovered suitable change periods by controlling the parameters, and our approach demonstrated good performance in simulations.

The purpose of this study is to explore how differences in parental investment in children's education influence the student's academic performance. In Section 2, this study adopts fuzzy theory to develop a methodology to represent the linear relationship between parental educational investment and student academic performance. Section 3 presents the findings of this study on the effect of educational investment on academic performance. Section 4 concludes the work with future directions.

2. Methods.

2.1. Fuzzy statistical analysis. The detection of trends and the measurement of change points is an interesting topic in copula analysis. Before attempting to identify change points or districts, several fundamental questions must be addressed. What does a change point or district mean? Can a clear definition of a change point or district be arrived at? How can change points be determined if the structure for a variable changes gradually? These problems concerning semantic interpretation and fuzzy statistical analysis have long bothered many researchers. For this reason, Zadeh proposed fuzzy set theory – a generalization of classical set theory – to accommodate semantic and conceptual fuzziness in statements [12]. Fuzzy theory intrinsically pertains to linguistic variables, a characteristic that can help researchers represent uncertainty in social phenomena.

In this study, we use fuzzy logic to address change periods and trend problems in correlation analysis. However, these detection techniques are based on the assumption that the underlying population exhibits the characteristic of having a significant change point [13]. Using fuzzy set theory, Wu and Chen proposed a procedure for detecting the change period in a nonlinear time series [13]. Nevertheless, when working with correlations with switching regimes, in addition to change point detection, the properties of change periods should also be considered.

In this study, we quantitatively analyzed survey data. A fuzzy questionnaire better captures the intrinsic uncertainty of human thought relative to its traditional counterpart, allowing the respondent to provide answers that accord more with what and how they think through the concepts of fuzzy set membership and interval values. Leveraging fuzzy statistical analysis and soft computing methods, fuzzy theory is a discipline concerning the quantitative representation of the language underpinning human thought.

2.2. Fuzzy correlation coefficient. Let X_i, Y_i be two sequence of a paired fuzzy data on a population Ω, and $X_i = [a_i, b_i]$, $Y_i = [c_i, d_i]$ $(i = 1, 2, \ldots, n)$, with its pair of centroids being (cx_i, cy_i) and pair of areas being $\|x_i\| = area(x_i)$, $\|y_i\| = area(y_i)$.

$$cr_{xy} = \frac{\sum_{i=1}^{n}\left(cx_i - \overline{cx}\right)\left(cy_i - \overline{cy}\right)}{\sqrt{\sum_{i=1}^{n}\left(cx_i - \overline{cx}\right)^2}\sqrt{\sum_{i=1}^{n}\left(cy_i - \overline{cy}\right)^2}} \tag{1}$$

$$\lambda ar_{xy} = 1 - \frac{\ln\left(1 + |ar_{xy}|\right)}{|ar_{xy}|}$$

where

$$ar_{xy} = \frac{\sum_{i=1}^{n} \left(\|x_i\| - \|\overline{x}_i\|\right)\left(\|y_i\| - \|\overline{y}_i\|\right)}{\sqrt{\sum_{i=1}^{n} \left(\|x_i\| - \|\overline{x}_i\|\right)^2}\sqrt{\sum_{i=1}^{n} \left(\|y_i\| - \|\overline{y}_i\|\right)^2}}. \quad (2)$$

Then, fuzzy correlation is defined as follows.

(i) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, \min[1, cr_{xy} + \lambda ar_{xy}])$.
(ii) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (cr_{xy} - \lambda ar_{xy}, cr_{xy})$.
(iii) When $cr_{xy} < 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, cr_{xy} + \lambda ar_{xy})$.
(iv) When $cr_{xy} < 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (\max[-1, c_{xy} - \lambda ar_{xy}], cr_{xy})$.

However, if one of the terms is a constant, irrespective of what the other variables do to the (x, y) pair, then λar_{xy} is zero; that is, the fuzzy correlation coefficient and Pearson's correlation coefficient are equal.

Example 2.1. *Suppose we have the following data presented in Table 1.*

TABLE 1. Numerical example of interval-valued fuzzy data

Sample	X			Y		
	Data	Centroid	Area (length)	Data	Centroid	Area (length)
A	[23, 25]	24	2	[1, 2]	1.5	1
B	[21, 26]	23.5	5	[2, 3]	1.5	1
C	[29, 35]	27	6	[0, 1]	0.5	1
D	[28, 30]	29	2	[2, 4]	2	2
E	[25, 34]	29.5	9	[2, 5]	3.5	3
(Fuzzy) mean		26.6	4.8		1.8	1.6

In that case, the correlation between the two centroids is as follows:

$$cr_{xy} = \frac{\sum_{i=1}^{n}(cx_i - 26.6)(cy_i - 1.8)}{\sqrt{\sum_{i=1}^{n} (cx_i - 26.6)^2}\sqrt{\sum_{i=1}^{n} (cy_i - 1.8)^2}} = 0.54.$$

Similarly, the correlation between the two lengths is as follows:

$$ar_{xy} = \frac{\sum_{i=1}^{n} \left(\|x_i\| - 4.8\right)\left(\|y_i\| - 1.6\right)}{\sqrt{\sum_{i=1}^{n} \left(\|x_i\| - 4.8\right)^2}\sqrt{\sum_{i=1}^{n} \left(\|y_i\| - 1.6\right)^2}} = 0.53.$$

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + 0.53)}{0.53} = 0.19.$$

Because the centroid correlation is $ar_{xy} \geq 0$ and the area (length) correlation is $\lambda ar_{xy} \geq 0$, the fuzzy correlation is $(r, \min[1, r + \lambda ar_{xy}]) = (0.54, \min[1, 0.72]) = (0.54, 0.72)$. This implies that the relationship between X and Y is quite high.

2.3. **Fuzzy hierarchy of correlation.** Human thinking is complex, uncertain, and sometimes surprising. However, the statistical relationships obtained from traditional questionnaires come only in the forms of positive correlations, negative correlations, or no correlation. Moreover, analyses of correlations in crisp (i.e., traditional, non-fuzzy) data sets require such data sets to be large and normally distributed, with Pearson's correlation coefficient typically used to calculate the R value. Therefore, in this study, we further explored the use of the fuzzy correlation coefficient to represent the linear relationship between two variables. We take the median of a data set as a comparative reference point to divide the linear relationship into two districts. This method is called the fuzzy hierarchy of correlation (FHC) method.

For example, in a traditional study on a drug's effect, pharmacodynamic dosing may initially exhibit a good effect (illustrated in segment A of Figure 1) but the overall segment (segment A and segment B) indicates no linear relationship. If Figure 1 is divided into the two hierarchies of A and B, we find that the dosages and effects are not positively correlated, even after overdose causes by the effect of variation (segment B), thereby exhibiting a negative correlation. In summary, the original linear relationship should have zero correlations from the hierarchical relationship revealed by the analysis, which clearly shows the trend and the steering of the associated linear relationship.

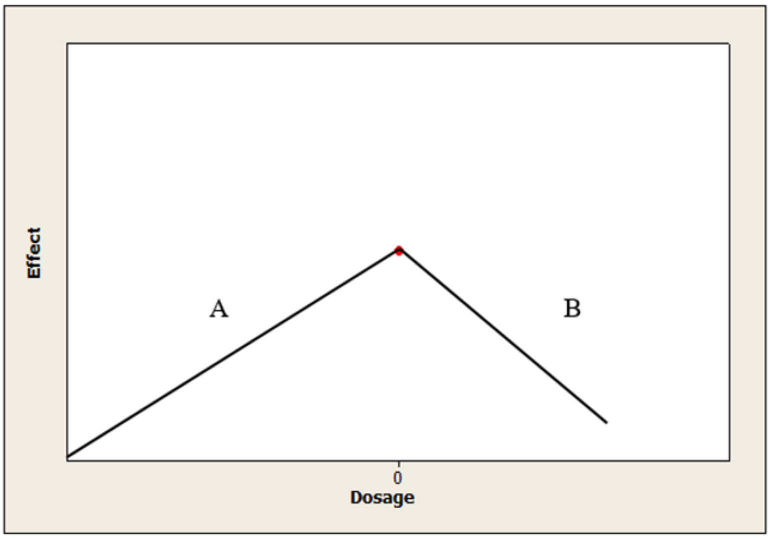

FIGURE 1. Dosage and effect of the linear relationship

2.4. **Decision of a switch regime.** We use the K-means method to detect an appropriate switch regime.

Step 1. Defuzzification for x in the fuzzy data pair (x, y).

Step 2. Ordering the fuzzy data pair (x, y) by the rank of x.

Step 3. Using the fuzzy K-means approach to classify the sequence of data (x, y).

Step 4. Calculating the correlation according to its regime.

3. **Results.** This study explored the correlation between Parental educational Investment (PI) and Student academic Performance (SP). However, many factors influence SP, such as school size, the urban-rural gap, and family socioeconomic background. Therefore, we adopt purposive sampling to recruit third-grade students of a junior high school (School A) in Taiwan to minimize the influence of other factors on the results. In addition, the factor of family socioeconomic background was similar for all students. Their parents were middle class, working as teachers or civil servants.

The following sections describe the use of the Spearman rank correlation test, fuzzy correlation coefficient, and FHC in our study.

3.1. **Spearman rank correlation test.** To detect whether PI is related to SP, this study adopted the Spearman rank correlation test.

Step 1. Defuzzify x in the fuzzy data (x, y).

Step 2. Use the defuzzification rule in continuous-interval fuzzy data.

Use the defuzzification rule $\{x_i = [a_i, b_i], i = 1, \ldots, n\}$ in a fuzzy data set as follows.

Let cf_i be the continuous defuzzification value of sample i, then $cf_i = c_i + \frac{l}{2\ln(e+c_i)}$, $l = b_i - a_i$, $c_i = \frac{1}{2}(b_i - a_i)$, where e is the base of the natural logarithm.

Step 3. Analyze the continuous defuzzification values (41 data points of PI are shown in Table 2).

Table 2. Defuzzification for data on parental investments

Sample	A_1	A_2	$\cdots$	A_{40}	A_{41}
Investments	$[0, 0]$	$[0, 0]$	$\cdots$	$[36500, 43500]$	$[43200, 56800]$
Defuzzification	0	0	$\cdots$	49305.6	50628.5

Step 4. Ranks of the defuzzification values of PI and SP are presented in Table 3.

Table 3. Rank of the variables of PI and SP

Sample	A_1	A_2	$\cdots$	A_{40}	A_{41}
PI rank	1	1	$\cdots$	40	41
SP rank	2	1	$\cdots$	9	39

Our sample comprised > 30 individuals. We used the central limit theorem to compute a Z value of 3.08. The result was significant, and the null hypothesis (H_0) was rejected. Therefore, PI and SP were found to be positively correlated.

3.2. **Fuzzy correlation coefficient.** As indicated in Table 4, the correlation between the centroids of PI and SP is as follows:

$$cr_{total(PI\bullet SP)} = \frac{\sum_{i=1}^{n}(cx_i - 15262.19)(cy_i - 344.73)}{\sqrt{\sum_{i=1}^{n}(cx_i - 15262.19)^2}\sqrt{\sum_{i=1}^{n}(cy_i - 344.73)^2}} = 0.33$$

If one of the terms is a constant, irrespective of what the other variable does for the (x, y) pair, the correlation ($\lambda ar_{PI\bullet SP}$) is zero.

Table 4. PI and SP interval-valued fuzzy data

Sample	Parents' educational investment (PI)			Students' academic performance (SP)		
	Data (x_i)	Centroid (cx_i)	Length	Data (y_i)	Centroid (cy_i)	Length
A_1	$[0, 0]$	0	0	230	230	0
A_2	$[0, 0]$	0	0	200	200	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_{40}	$[36500, 43500]$	40000	3500	332	332	0
A_{41}	$[43200, 56800]$	50000	6800	391	391	0
(Fuzzy) mean		15262.19	2034.87	344.73	344.73	0

Using Formula (1) on the data for 41 students, we computed the total correlation coefficient of PI and SP to be $cr_{total(PI\bullet SP)} = 0.33$, indicating a moderate correlation. Subsequently, we used Minitab 16.0 to draw a linear plot of the two variables (Figure 2).

3.3. **Fuzzy hierarchy of correlation.** First, the median interval value of $[7000, 13000]$ and the two interval values of $[4500, 7500]$ and $[11500, 14500]$ fell in the correlation transition district. To make the correlation coefficient more accurate, we deduced those two values that fell within the correlation transition district. Second, the linear plot was divided into three levels in a hierarchy, namely the first district, second district, and correlation transition district (Figure 3). Both the first and second districts contained data on 19 participants. Furthermore, the second district was deduced from the first district and correlation transition district.

As explained, the data of the first district were $[0, 0], [0, 0], \ldots, [4000, 6000]$ and the data of the second district were $[13600, 16400], [13500, 16500], \ldots, [43200, 56800]$. The correlation between the first district and the centroids of PI and SP is as follows:

$$cr_{first(PI\bullet SP)} = \frac{\sum_{i=1}^{n}(cx_i - 2223.68)(cy_i - 326.26)}{\sqrt{\sum_{i=1}^{n}(cx_i - 2223.68)^2}\sqrt{\sum_{i=1}^{n}(cy_i - 326.26)^2}} = 0.61$$

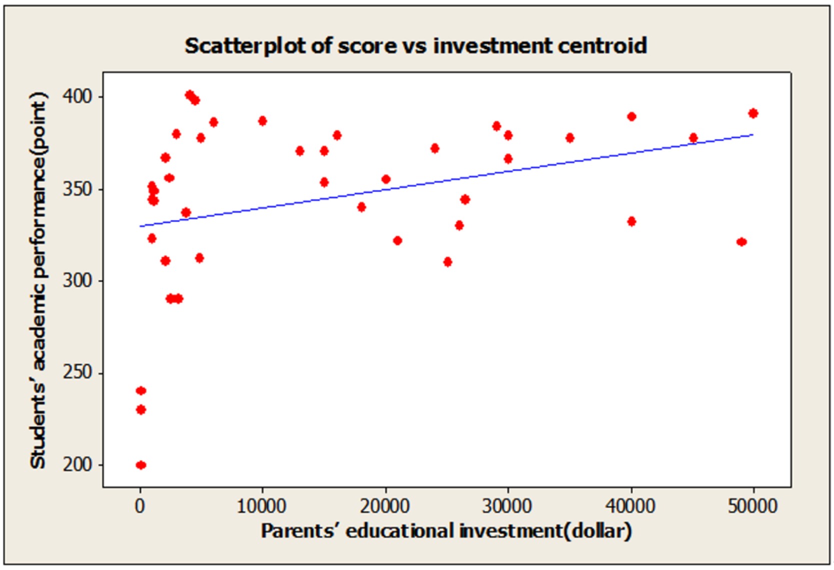

Figure 2. Linear plot of PI and SP

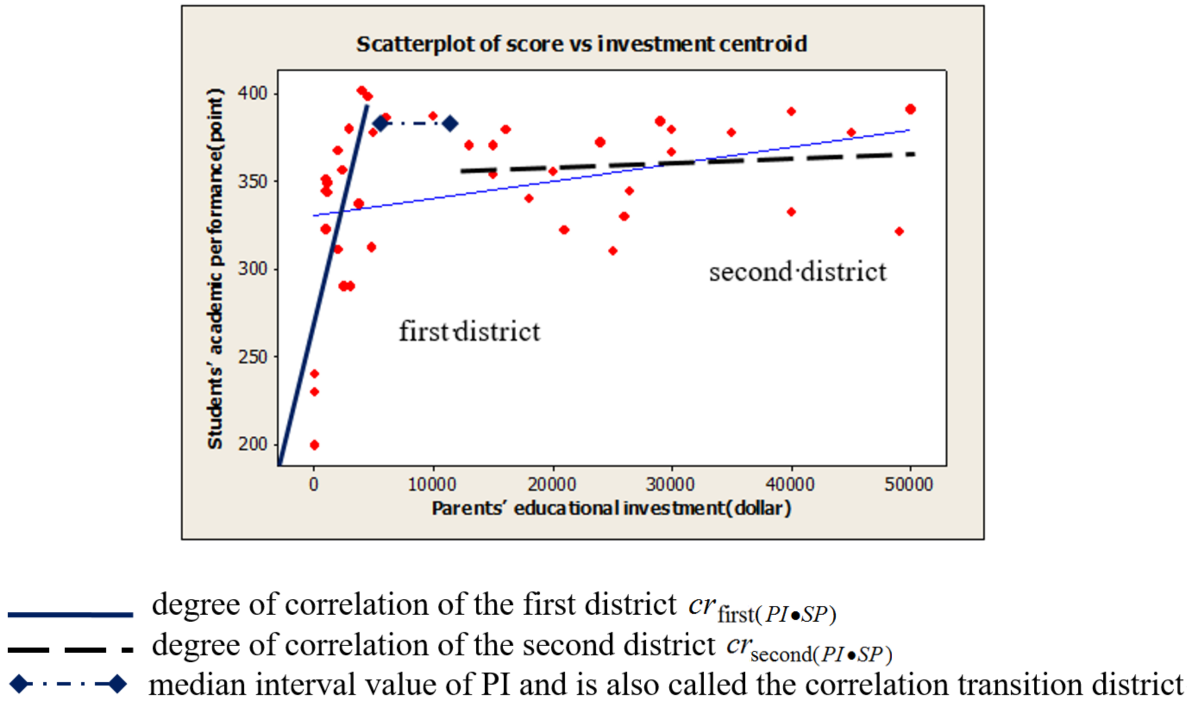

Figure 3. Linear plot of fuzzy hierarchy of correlation

If one of the terms is a constant, irrespective of what the other variables do for the (x, y) pair, the correlation $(\lambda ar_{PI\bullet SP})$ is zero.

Using Formula (2) on the data for 19 students, the results are shown in Table 5, we computed the correlation coefficient of PI and SP to be $cr_{first(PI\bullet SP)} = 0.61$, indicating a high correlation. Subsequently, we used Minitab 16.0 to plot the linear relationship for the first district (Figure 3). The results of this study agreed the finding of previous studies [14-16]. The parental educational investment (PI) and student academic performance (SP) are positively correlated. Besides, this study further identifies the existence of the correlation transition district. However, other factors also import for student academic performance, as parental involvement [16,17].

Table 5. First district of interval-valued fuzzy data

Sample of first district	Parents' educational investment (PI)			Students' academic performance (SP)		
	Data (x_i)	Centroid (cx_i)	Length	Data (y_i)	Centroid (cy_i)	Length
A_1	$[0, 0]$	0	0	230	230	0
A_2	$[0, 0]$	0	0	200	200	0
...	...	...	...	...	...	...
A_{18}	$[4600, 5000]$	4800	200	312	312	0
A_{19}	$[4000, 6000]$	5000	1000	377	377	0
(Fuzzy) mean		2223.68	217.36	326.26	326.26	0

Table 6. Second district of interval-valued fuzzy data

Sample of second district	Parents' educational investment (PI)			Students' academic performance (SP)		
	Data (x_i)	Centroid (cx_i)	Length	Data (y_i)	Centroid (cy_i)	Length
A_{21}	$[13600, 16400]$	15000	1400	354	354	0
A_{22}	$[13500, 16500]$	15000	1500	370	370	0
...	...	...	...	...	...	...
A_{40}	$[36500, 43500]$	40000	3500	332	332	0
A_{41}	$[43200, 56800]$	50000	6800	391	391	0
(Fuzzy) mean		29184.21	3857.89	357.47	357.47	0

4. **Conclusion.** We used fuzzy statistics to better represent the correlation between PI and SP in the form of the fuzzy hierarchy. Our sample comprised 41 students who had just graduated from junior high school (School A). The data on SP were the participants' scores in Taiwan's Basic Competence Test, and we compared the relationship of SP with the fuzzy interval functions of PI.

We suggest the following based on the findings.

1) $cr_{total(PI \bullet SP)} = 0.33$; PI and SP are positively correlated. Thus, parents should invest more in their children's education to improve academic performance.

2) Although PI and SP are positively correlated, this correlation does not entail that greater PI causes an increase in SP. Interpretations of our findings should not stretch beyond what the evidence warrants. We thus proposed FHC to avoid the pitfalls of excessive interpretation, using which we obtained the following correlation transition district.

3) $cr_{first(PI \bullet SP)} = 0.61 > cr_{total(PI \bullet SP)} = 0.33$. The correlation coefficient before the correlation transition district ($cr_{first(PI \bullet SP)}$) is higher than the total correlation coefficient ($cr_{first(PI \bullet SP)}$).

4) $cr_{second(PI \bullet SP)} = 0.16 < cr_{total(PI \bullet SP)} = 0.33$. The correlation coefficient after the correlation transition district ($cr_{first(PI \bullet SP)}$) is lower than the total correlation coefficient ($cr_{first(PI \bullet SP)}$). This implies that the relationship between PI and SP is not a causal one.

5) We used the FHC method to identify the existence of the correlation transition district. FHC was used to examine PI and SP to obtain the point and period of change. Our approach is also novel. In the future, we will apply the fuzzy polynomial regression model to analyze the trend and change cycle of the educational investment to academic achievement.

REFERENCES

[1] Organization for Economic Co-operation and Development, *PISA 2018 Results (Volume II): Where All Students Can Succeed*, DOI: 10.1787/b5fd1b8f-en, 2019.

[2] M. D. R. Evans, J. Kelley, J. Sikora and D. J. Treiman, Family scholarly culture and educational success: Books and schooling in 27 nations, *Research in Social Stratification and Mobility*, vol.28, no.2, pp.171-197, 2010.

[3] S. Thomson, Achievement at school and socioeconomic background: An educational perspective, *npj Science of Learning*, vol.3, no.5, pp.1-2, DOI: 10.1038/s41539-018-0022-0, 2018.

[4] M. Uzair-ul-Hassan, M. S. Farooq, M. P. Akhtar and I. Parveen, Teachers' politeness as a predictor of students' self-esteem and academic performance, *Bulletin of Education and Research*, vol.39, no.1, pp.229-243, 2017.

[5] K. Laidra, H. Pullmann and J. Allik, Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school, *Personality and Individual Differences*, vol.42, pp.441-451, 2007.

[6] F. L. Van Voorhis, Interactive science homework: An experiment in home and school connections, *National Association of Secondary School Principals Bulletin*, vol.85, pp.20-32, 2001.

[7] E. M. Horvat, E. B. Weininger and A. Lareau, From social ties to social capital: Class differences in the relations between schools and parent networks, *American Educational Research Journal*, vol.40, no.2, pp.319-350, 2003.

[8] J. Liu, P. Peng and L. Luo, The relation between family socioeconomic status and academic achievement in China: A meta-analysis, *Educational Psychology Review*, vol.32, pp.49-76, DOI: 10.1007/s10648-019-09494-0, 2020.

[9] A. Lareau, *Home Advantage: Social Class and Parental Intervention in Elementary Education*, 2nd Edition, Rowman & Littlefield, Lanham, MD, 2000.

[10] B. Wu, W. Lai, C. L. Wu and T. K. Tienliu, Correlation with fuzzy data and its applications in the 12-year compulsory education in Taiwan, *Communications in Statistics – Simulation and Computation*, vol.45, no.4, pp.1337-1354, DOI: 10.1080/03610918.2013.827712, 2016.

[11] Y. Yang, J. Ruan, J. Li, B. Liu and X. Kong, Route choice behavior model with guidance information based on fuzzy analytic hierarchy process, *International Journal of Innovative Computing, Information and Control*, vol.14, no.1, pp.363-370, 2018.

[12] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol.8, pp.338-353, 1965.

[13] B. Wu and M. Chen, Use of fuzzy statistical methods in change periods detection, *Applied Mathematics and Computation*, vol.99, pp.241-254, 1999.

[14] H. Chou, Verification of coleman's social capital theory in Taiwan area: The relationships of the family, social capital and academic achievement, *Contemporary Educational Research Quarterly*, vol.14, no.4, pp.1-28, 2006.

[15] J. Hsieh, M. Chen and M. Tsai, Examining the impact of demographic factors on low-achievers' Chinese growth trajectories, *National Chiayi University Journal of the Educational Research*, vol.42, pp.27-65, 2019.

[16] V. Gubbins and G. Otero, Parental involvement and low-SES children's academic achievement in early elementary school: New evidence from Chile, *Educational Studies*, vol.46, no.1, pp.1-22, DOI: 10.1080/03055698.2019.1620691, 2019.

[17] S. Park and S. D. Holloway, The effects of school-based parental involvement on academic achievement at the child and elementary school level: A longitudinal study, *The Journal of Educational Research*, vol.110, no.1, pp.1-16, DOI: 10.1080/00220671.2015.1016600, 2017.