

Factors Affecting Whether College Students Go Back to School after Suspension

Chun-Min Hung^{1*}, Chun-Yen Chung², Yan-Kuin Su³, Chin-Chiuan Lin⁴

¹(Department Of Information Management, Kun Shan University, Taiwan)

²(Department Of Information Engineering, Kun Shan University, Taiwan)

³(President, Kun Shan University, Taiwan)

⁴(Department Of Business Administration, Kun Shan University, Taiwan)

Abstract : The present study investigated factors that might affect college students going back to school after being suspended in Taiwan based on the institutional research data of KS University. Results showed that the grades, enrolled department, entrance method, academic performance and the interaction between grades and enrolled department are the factors that significantly affect whether college students go back to school after being suspended in Taiwan. Juniors were the college students with the lowest ratio of those going back to school after being suspended. College of applied human ecology (department 2) showed the lowest ratio of students going back to school compared to other departments. Students who applied through direct application showed a higher ratio of going back to school than those entering based on the unified exam. Students with lower academic performance had a higher ratio of going back to school. Interaction effect also indicated that students with the lowest ratio of going back to school were juniors from College of applied human ecology.

Keywords - Factors; Back to school; Suspension; Linear information model; Institutional research

I. Introduction

Student suspension or withdrawal has become an import issue for private colleges in Taiwan. Fig. 1 shows the flow chart of student suspension and withdrawal and the results in Taiwan.

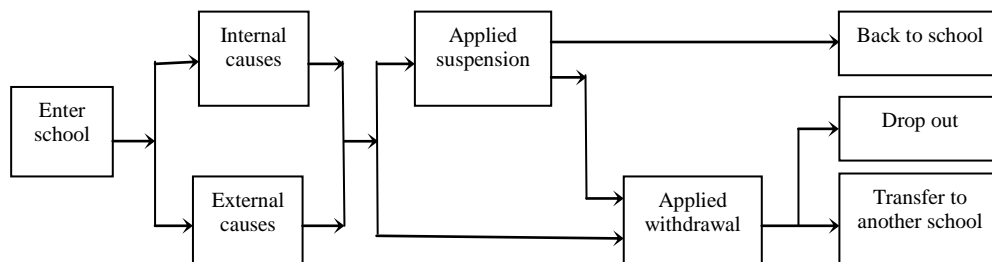


Fig. 1. Flow chart for student suspension and withdrawal and the results

To better understand the circumstances, Hung et al. [1] constructed a conceptual model (Fig. 2) based on a literature review to reveal the real (internal) causes of suspension and withdrawal of college students in Taiwan.

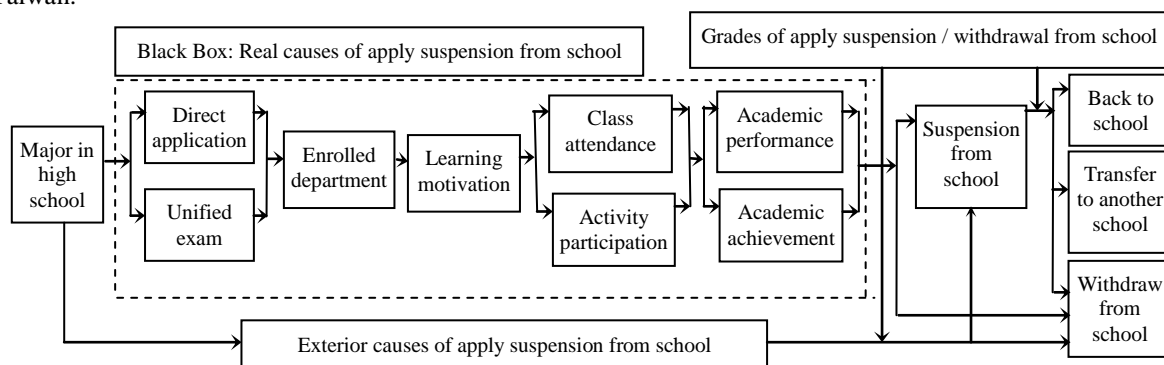


Fig. 2. Model of suspension and withdrawal from school in Taiwan

Hung et al. [2] verified the conceptual model which they proposed based on institutional research (IR) data and indicated that the conceptual model (Fig. 3) can actually reflect the current situation in Taiwan.

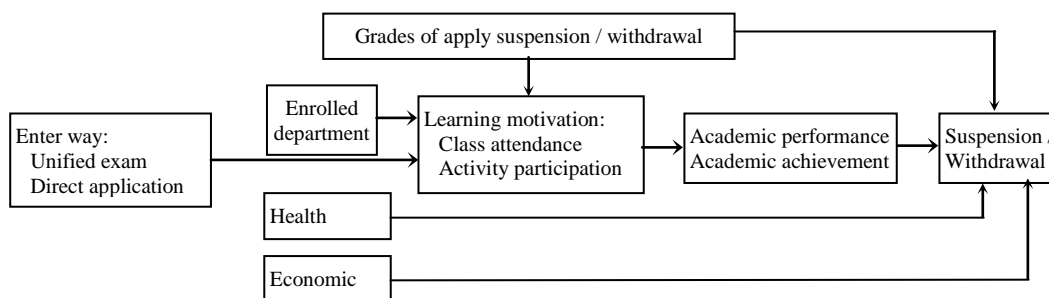


Fig. 3. Significant path coefficients for suspension and withdrawal

However, their results only indicate the significance coefficients of the path between the factors rather than recognizing the impact of each individual factor on suspension or withdrawal. Therefore, Hung et al. [3] employed a logistic regression (LR) method associated with information geometry to analyze the data. Results showed that the class attendance and interaction between class attendance and academic performance significantly affect college students’ suspension and withdrawal in Taiwan.

As a result, if the students can go back to school after being suspended, then the impact of suspension is slight. Therefore, to understand the factors that affect whether college students go back to school after being suspended is also an important issue. Thus, the present study investigated the factors related to this issue.

The basic theory of the geometrically supported LR analysis method (linear information model, LIM) [4-7], which characterizes the geometry of the association between categorical variables, was introduced by Hung et al. [3]. Therefore, the present study directly analyzes IR data by employing LIM.

II. Practical Data Analysis

An IR was conducted to evaluate the status of college students in KS University who were suspended from school from 2010-2014. We examined how various factors affected predictions of whether students would go back to school after being suspended by using a LIM model.

Based on data likelihood decomposition, an information approach supported by geometry theory for selecting the main and interaction effects of the predicting variables is introduced to LR analysis.

4.1. Data And Codes

In the IR data of KS University, a nominal variable is used to define the status of the response variable (back to school=1 and withdrawal=2), denoted by R=1 and 2. Seven predictors, each coded as “1 to 2 or 1 to 5” are used. Thus, the data consist of a multivariate contingency table of seven variables, having 800 cells and a total count of 1,657. Table I lists the codes of the seven prediction variables and the response variable, and their descriptions.

Table I. List of codes and descriptions of the variables

Variable	Code	Description
Entrance method	EM, x_1	1: direct application; 2: unified exam.
Living city	LC, x_2	1: Tainan city; 2: other cities.
Gender	GE, x_3	1: Male; 2: Female.
Enrolled department	ED, x_4	1: college of creative media; 2: college of applied human ecology; 3: college of information technology; 4: college of business and management; 5: college of engineering.
Class attendance	CA, x_5	1: number of class absence < 5; 2: others.
Academic performance	AP, x_6	1: average scale > 60; 2: others.
Grade	GR, x_7	1: applied suspension at first grade; 2: applied suspension at second grade; 3: applied suspension at third grade; 4: applied suspension at fourth grade; 5: applied suspension after fourth grade.
Result	R	1: back to school; 2: withdrawal.

4.2. Classical LR Analysis

Table II shows the association between R (result) and each individual prediction variable.

Table II. Association between R and each individual factor

Variable	Code	Result		Ratio of going back to school
		back to school	withdrawal	
EM, x_1	1	686	60	91.96%
	2	747	164	82.00%
LC, x_2	1	511	62	89.18%
	2	922	162	85.06%
GE, x_3	1	911	130	87.51%
	2	522	94	84.74%
ED, x_4	1	277	27	91.12%
	2	169	53	76.13%
	3	230	36	86.47%
	4	295	50	85.51%
	5	462	58	88.85%
CA, x_5	1	673	128	84.02%
	2	760	96	88.79%
AP, x_6	1	1062	196	84.42%
	2	371	28	92.98%
GR, x_7	1	178	17	91.28%
	2	227	24	90.44%
	3	223	77	74.33%
	4	415	47	89.85%
	5	389	59	86.83%

A full model for the case of using seven predictors $\{x_1, x_2, \dots, x_7\}$ for the result R is:

$$\begin{aligned}
 &I(\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, R) \\
 &= 7 x_i \\
 &+ 21 x_i x_j \\
 &+ 35 x_i x_j x_k \\
 &+ 35 x_i x_j x_k x_l \\
 &+ 21 x_i x_j x_k x_l x_m \\
 &+ 7 x_i x_j x_k x_l x_m x_n \\
 &+ 1 x_i x_j x_k x_l x_m x_n x_o \dots \dots \dots (1)
 \end{aligned}$$

The full model of classical LR analysis for equation (1) would include 7 main effects, 21 two-order interactions, 35 three-order interactions, 35 four-order interactions, 21 five-order interactions, 7 six-order interactions, and 1 seven-order interaction. It would take a very long time to obtain the full model results; furthermore, the full model results would be too complex to interpret.

If we select the factors one by one, it would still be too complex to calculate the relationship between the seven prediction variables and the response variable because any variable of the seven prediction variables can be coded as x_1, x_2, \dots , and x_7 . In this case, there would be 5,040 (7!) combinations that need to be calculated. Thus, selecting variables efficiently for the LR model is clearly the major problem.

Thus, classical LR analysis methods usually assume that the high-order interactions are insignificant. However, this approach might neglect some significant high-order interactions and lead to incorrect interpretation.

Therefore, reducing items without missing significant interactions is very important, especially for significant high-order interactions.

4.3. LIM Analysis

A basic approach is to eliminate redundant predictors and to test LR models that can be interpreted using the least number of significant interaction terms. A straightforward extension of the MI identities in (1) to high-way tables is examined for IR data analysis in this study. An extension of the first equation of (1) to the case of using seven predictors $\{x_1, x_2, \dots, x_7\}$ for the response variable R is:

$$\begin{aligned}
 &I(\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, R) \\
 &= I(R, x_1/x_7, x_6, x_5, x_4, x_3, x_2) \\
 &+ I(R, x_2/x_7, x_6, x_5, x_4, x_3) \\
 &+ I(R, x_3/x_7, x_6, x_5, x_4) \\
 &+ I(R, x_4/x_7, x_6, x_5) \\
 &+ I(R, x_5/x_7, x_6) \\
 &+ I(R, x_6/x_7)
 \end{aligned}$$

$$+ I(R,x_7).....(2)$$

Identity (2) is constructed by the rule of selecting the first least significant (7th order) conditional MI (CMI) term, then selecting the least significant 6th order CMI term, and continuing until reaching the last 2nd order CMI term $I(R,x_6/x_7)$. Table III shows the calculation of the decomposition of identity (2) for each variable.

Table III. Calculation of the decomposition of identity (2)

Variable	$I(R,x_i/x_j,...)$			$Int(R,x_i,x_j,...)$			$I(R,x_i/x_j,...)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
EM, x_1	249.132	400	1.000	209.988	399	1.000	39.144	1	0.000
LC, x_2	107.578	400	1.000	102.479	399	1.000	5.099	1	0.024
GE, x_3	122.624	400	1.000	116.319	399	1.000	6.305	1	0.012
ED, x_4	290.952	640	1.000	274.615	636	1.000	16.337	4	0.003
CA, x_5	130.776	400	1.000	130.718	399	1.000	0.058	1	0.810
AP, x_6	85.923	400	1.000	70.406	399	1.000	15.517	1	0.000
GR, x_7	350.337	640	1.000	293.222	636	1.000	57.115	4	0.000

The efficient way to select an initial sequence of variables is according to the significance level of the main and interaction term based on MI and CMI. Therefore, identity (2) is updated as:

$$\begin{aligned}
 &I(R, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}) \\
 &= Int(R, x_1, \{x_2, x_3, x_4, x_5, x_6, x_7\}) \\
 &+ I(R, x_1 // x_2, x_3, x_4, x_5, x_6, x_7) \\
 &+ Int(R, x_2, \{x_3, x_4, x_5, x_6, x_7\}) \\
 &+ I(R, x_2 // \{x_3, x_4, x_5, x_6, x_7\}) \\
 &+ Int(R, x_3, \{x_4, x_5, x_6, x_7\}) \\
 &+ I(R, x_3 // \{x_4, x_5, x_6, x_7\}) \\
 &+ Int(R, x_4, \{x_5, x_6, x_7\}) \\
 &+ I(R, x_4 // \{x_5, x_6, x_7\}) \\
 &+ Int(R, x_5, \{x_6, x_7\}) \\
 &+ I(R, x_5 // \{x_6, x_7\}) \\
 &+ Int(R, x_6, \{x_7\}) \\
 &+ I(R, x_6 // \{x_7\}) \\
 &+ Int(R, x_7).....(3)
 \end{aligned}$$

4.4. Determine Initial Sequence Of Factors

Table III showed factor GR is the most significant factor. Therefore, the GR factor is first entered into the LR model. Then repeating the calculation procedure (Appendix Table I-VI), the factor ED is entered into the LR model secondly, followed by EM, AP, CA, LC and GE. When the entry sequence of the variables is determined, the next problem is the selection of a proper LR model.

Table IV. Sequential decomposed CMI components of identity (2)

MI, CMI Terms	$I(R, X_{(i)} / R^{(i)} X_{(i)})$			$Int(R, X_{(i)} / R^{(i)} X_{(i)})$			$I(R, X_{(i)} // R^{(i)} X_{(i)})$		
	CMI	df	Sig.	Interaction	df	Sig.	Partial Asso.	df	Sig.
$I(R, x_7/x_1, x_2, x_3, x_4, x_5, x_6)$	350.337	640	1.000	293.222	636	1.000	57.115	4	0.000
$I(R, x_4/x_1, x_2, x_3, x_5, x_6)$	175.374	128	0.003	147.474	124	0.074	27.900	4	0.000
$I(R, x_1/x_2, x_3, x_5, x_6)$	72.643	16	0.000	42.703	15	0.000	29.940	1	0.000
$I(R, x_6/x_2, x_3, x_5)$	34.666	16	0.004	17.302	15	0.301	17.364	1	0.000
$I(R, x_5/x_2, x_3)$	9.130	4	0.058	2.595	3	0.458	6.535	1	0.011
$I(R, x_2/x_3)$	7.281	2	0.026	1.668	1	0.197	5.613	1	0.018
$I(R, x_3)$	2.510	1	0.113	-	-	-	-	-	-

4.5. Selection Of A Proper LR Model

An LR model is constructed from identity (2) using the hierarchical set of variable parameters $\{x_7, x_4, x_7x_4, x_1, (x_7x_4)x_1, x_6, (x_7x_4x_1)x_6, x_5, (x_7x_4x_1x_6)x_5, x_2, (x_7x_4x_1x_6x_5)x_2, x_3\}$ as identity (4).

$$\begin{aligned}
 &Logit(R) \{x_6, x_7, x_5, x_1, x_3, x_4, x_2\} \\
 &= 1.886^* \\
 &- 0.823 (x_7=3)^* \\
 &- 0.747 (x_4=2)^* \\
 &+ 2.285 (x_7=3, x_4=1)^* \\
 &- 1.055 (x_1=1)^*
 \end{aligned}$$

$$\begin{aligned}
 & - 1.538 x_7x_4x_1 \\
 & - 0.970 (x_{6=1})^* \\
 & - 13.401 x_7x_4x_1x_6 \\
 & - 0.083 x_5 \\
 & + 13.824 x_7x_4x_1x_6x_5 \\
 & + 0.419 x_2 \\
 & - 14.280 x_7x_4x_1x_6x_5x_2 \\
 & + 0.644 x_3 \dots\dots\dots(4)
 \end{aligned}$$

III. Results

Analysis of results shows only the main effect of x_7, x_4, x_1, x_6 and interaction of x_7x_4 reached a statistical level of significance ($p < 0.01$), indicating that grades, enrolled department, entrance method, academic performance and interaction between grades and enrolled department significantly affect the ratio students going back to school after being suspended.

Grades (x_7) significantly affect the ratio students going back to school after suspension. Table V shows juniors ($x_{7=3}$) had the lowest ratio of students going back to school. In contrast, freshman ($x_{7=1}$) showed the highest ratio of students going back to school.

Table V. Effect of grades on ratio of going back to school after suspension

Grade	Result	Number of students	Ratio of going back to school
1	back to school	178	91.28%
	withdrawal	17	
2	back to school	227	90.44%
	withdrawal	24	
3	back to school	223	74.33%
	withdrawal	77	
4	back to school	416	89.85%
	withdrawal	47	
5	back to school	389	86.83%
	withdrawal	59	

Enrolled department (x_4) significantly affects the ratio of students going back to school after suspension. Table VI shows department 2 resulted in the lowest ratio of students going back to school. In contrast, department 1 had the highest ratio of students going back to school.

Table VI. Effect of enrolled department on ratio of going back to school after suspension

Enrolled department	Result	Number of students	Ratio of going back to school
1	back to school	277	91.12%
	withdrawal	27	
2	back to school	169	76.13%
	withdrawal	53	
3	back to school	230	86.47%
	withdrawal	36	
4	back to school	295	85.51%
	withdrawal	50	
5	back to school	462	88.85%
	withdrawal	58	

Entrance method (x_1) significantly affects the ratio of students going back to school after suspension. The students who entered the school by direct application ($x_{1=1}$) had a higher ratio of students going back to school than those who entered via the unified exam ($x_{1=2}$).

Table VII. Effect of entrance method on the ratio of going back to school after suspension

Entrance method	Result	Number of students	Ratio of going back to school
1	back to school	686	91.96%
	withdrawal	60	
2	back to school	747	82.00%
	withdrawal	164	

Academic performance (x_6) significantly affects the ratio of students going back to school after suspension. The students with better academic performance ($x_7=1$) had a lower ratio of going back to school than those with lower academic performance ($x_7=2$). This result comes into conflict with our exceptions and might result from it being harder for students with poor academic performance students to transfer to another school. Therefore, they tend to come back to school after being suspended when they fail to transfer.

Table VIII. Effect of academic performance on the ratio of going back to school after suspension

Academic performance	Result	Number of students	Ratio of going back to school
1	back to school	1062	84.42%
	withdrawal	196	
2	back to school	371	92.98%
	withdrawal	28	

The interaction effect between grades and enrolled department significantly affects the ratio of students going back to school after suspension. Table IX shows the association between grades and the enrolled department. Table IX indicates that college of applied human ecology (department 2) resulted in lowest ratio of junior students going back to school; in contrast, college of creative media (department 1) had the highest ratio of senior students going back to school.

Table IX. Association between grades and enrolled department with R

Grade	1	Result	Enrolled department				
			1	2	3	4	5
	1	back to school	40	21	22	39	56
		withdrawal	2	4	3	3	5
	2	back to school	54	30	53	46	44
		withdrawal	3	6	6	2	9
	3	back to school	51	44	25	46	57
		withdrawal	3	29	4	24	15
	4	back to school	64	41	62	77	172
		withdrawal	2	5	10	8	22
	5	back to school	68	33	68	87	133
		withdrawal	17	9	13	11	9

IV. Conclusions

Although there are many factors that might affect whether college students go back to school after suspension, however, the present study found that grades, enrolled department, entrance method, and academic performance significantly affect the ratio of students going back to school after being suspended. Furthermore, the interaction between grades and enrolled department also significantly affects the ratio of students going back to school after suspension.

References

- [1]. C.-M. Hung, C.-Y. Chung, Y.-K. Su and C.-C. Lin, Decision model of suspension or withdrawal of college students in Taiwan: Constructing a conceptual model, *British Journal of Education*, 4(3), 2016, pp. 89-98.
- [2]. C.-M. Hung, C.-Y. Chung, Y.-K. Su and C.-C. Lin, Decision model of suspension or withdrawal of college students in Taiwan: Verification of model, *British Journal of Education*, 4(4), 2016, pp. 65-74.
- [3]. C.-M. Hung, C.-Y. Chung, Y.-K. Su and C.-C. Lin, Factors affecting college students applied suspended or withdraw from school in Taiwan, *IOSR Journal of Research & Method in Education*, 6(3), 2016, pp. 1-7.
- [4]. P.E. Cheng, J.W. Liou, M. Liou and J.A.D. Aston, Linear information models: An introduction, *Journal of Data Science*, 5, 2007, pp. 297-313.
- [5]. P.E. Cheng, M. Liou, J.A.D. Aston and A.C. Tsai, (2008) Information identities and testing hypotheses: Power analysis for contingency tables. *Statistica Sinica*, 18, 2008, pp. 535-558.
- [6]. P.E. Cheng, M. Liou and J.A.D. Aston, Likelihood ratio tests with three-way tables, *Journal of the American Statistical Association*, 105, 2010, pp. 740-749.

- [7]. P.E. Cheng, J.W. Liou, M. Liou and J.A.D. Aston, Data information in contingency tables: A fallacy of hierarchical loglinear models, Journal of Data Science, 4, 2006, pp. 387-398.

Appendix Table I. Calculation of decomposed after deleted GR

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
EM, x_1	146.113	80	0.000	121.088	79	0.002	25.025	1	0.000
LC, x_2	64.869	80	0.990	58.964	79	0.955	5.905	1	0.015
GE, x_3	97.351	80	0.091	95.065	79	0.105	2.286	1	0.131
ED, x_4	175.374	128	0.003	147.474	124	0.074	27.900	4	0.000
CA, x_5	104.522	80	0.034	104.040	79	0.031	0.482	1	0.488
AP, x_6	60.832	80	0.946	44.593	79	0.999	16.239	1	0.000

Appendix Table II. Calculation of decomposed after deleted GR and ED

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
EM, x_1	72.643	16	0.000	42.703	15	0.000	29.940	1	0.000
LC, x_2	26.710	16	0.045	23.024	15	0.084	3.686	1	0.055
GE, x_3	41.234	16	0.001	31.844	15	0.007	9.390	1	0.002
CA, x_5	39.901	16	0.001	39.068	15	0.001	0.833	1	0.361
AP, x_6	34.666	16	0.004	17.302	15	0.301	17.364	1	0.000

Appendix Table III. Calculation of decomposed after deleted GR, ED and EM

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
LC, x_2	13.564	8	0.094	9.560	7	0.215	4.004	1	0.045
GE, x_3	20.145	8	0.010	14.450	7	0.044	5.695	1	0.017
CA, x_5	10.376	8	0.240	10.326	7	0.171	0.050	1	0.823
AP, x_6	34.019	8	0.000	13.368	7	0.064	20.651	1	0.000

Appendix Table IV. Calculation of decomposed after deleted GR, ED, EM and AP

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
LC, x_2	6.323	4	0.176	2.531	3	0.470	3.792	1	0.051
GE, x_3	5.611	4	0.230	2.446	3	0.485	3.165	1	0.075
CA, x_5	9.130	4	0.058	2.595	3	0.458	6.535	1	0.011

Appendix Table V. Calculation of decomposed after deleted GR, ED, EM, AP and CA

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
LC, x_2	7.281	2	0.026	1.668	1	0.197	5.613	1	0.018
GE, x_3	4.167	2	0.125	1.668	1	0.197	2.499	1	0.114

Appendix Table VI. Calculation of decomposed after deleted GR, ED, EM, AP, CA and LC

Factor	$I(R_i, x_j x_j, \dots)$			$Int(R_i, x_i, x_j, \dots)$			$I(R_i, x_i // x_i, \dots)$		
	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.	Likelihood ratio	df	Sig.
GE, x_3	2.510	1	0.113	-	-	-	-	-	-