2012 South Central SAS Users Group Annual Conference
Houston, Texas, November 5-6, 2012

Institutional Research Intelligence (IRI): Go Beyond Reporting

Harjanto Djunaidi[1]

Abstract

Recent dynamic changes in the economy, government funding regulations and education industry have forced many institutions to operate with a tight if not a smaller budget. Some schools have to freeze faculty's and staff's salary increase for many years in order to stay afloat. Others have to reduce students' services and at some points also lower the quality of delivered services to the students by increasing the class size. In order to cope and adapt with the new realities, colleges have no other choice, but to increase their operational efficiency. The *Institutional Research Intelligence* (IRI) is a new concept that provides tools to help administrators such as the President, VPs, Deans, Department Heads, Program Directors and other school decision makers to achieve the institution's goals more effectively. Given the significant and dynamic changes in the industry, universities and colleges in the US cannot just operate BAU (business as usual). Rather, any strategic planning and decisions need to be based on past data and data-driven information. IRI is the next needed concept which offers future tools to provide vital information to the administrators. IRI embeds new concepts, new ways of thinking with new mindsets to manage higher education institutions. This approach combines several disciplines such as statistical analyses, econometrics, market intelligence/research, competitive analyses, computer programming and the "old" IR in the decision making process. IRI helps institutions to achieve their goals while delivering the best value to their clientele.

[1] Updated 2017. Contact email: AAEA.US@yahoo.com. IRI-Analytics initiator/innovator. Visit http://www.aaea.us for more information on IRI/Education Analytics.

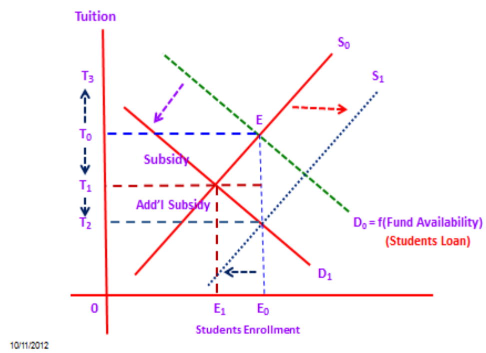
Introduction

Perhaps, only a handful people have expected the 2008 financial crises will ever happen in the US and the world. When the bankruptcy news of Lehman Brothers' broke out the financial market communities from Shanghai to Frankfurt were stunned. The chain reactions followed by the collapse of other big banks have forced the US government to intervene in order to prevent further chaos in the global financial market. Will the same crises happen in the education industry?

With the students' loan default on the rise, fewer jobs are available to absorb the college graduates, less availability of taxpayers' money to fund higher education, increasing excess supply of higher education services and the newly introduced SAP criteria on the government financial assistance (Title IV) are the ingredients for a perfect storm for those who operate under BAU mindset. As shown in Table 1 during the period of 2005 to 2010, the number of higher education institutions (HEIs) in the US has increased about 5.10%. Increasing number of HIEs has caused supply of offered HIEs' services to grow positively. Using a comparative static equilibrium analysis, one might be able to analyze the potential impacts of supply increase to the whole industry as shown in Figure 1. Increasing in education services will shift the supply curve outward from S_0 to S_1. Before anything else changes in the system, students are willing to pay the tuition at T_0, where D_0 and S_0 intersect.

After the supply has increased, the tuition ***is supposed*** to decrease from T_0 to T_1. However, in reality that never happens. In fact, in the past several years, college tuition on average has increased by $764.21 per year (*Source: Harjanto Djunaidi: Institutional Research Intelligence: Go beyond Reporting, upcoming book, summer, 2013*[2]).

Figure 1 – Decreasing Funding and Increasing Supply: Impact on State Subsidy



Instead of paying at T_0, students and their family have a pretty good chance to pay higher than T_0, for example at T_3. If one brings the impacts of decreasing students' loan and the potential effects of 2011 policy changes on Title IV financial assistance known as SAP[3] into the analyses, she or he will see that either additional subsidy or potential additional commercial loan are needed by the industry to keep student enrollment at E_0, where D_0 and S_0 intersect. If the amount of money that students can borrow is reduced significantly due to SAP new regulations then some of the borrowers

---

[2] The electronic version of the book is planned to be launched in the end of November, 2012 and it can be accessed through the following website: http://www.aaea.us/

[3] The government has introduced the SAP (Satisfactory Academic Progress) criteria concerning Title IV financial assistance. Section § 668.34 stated that "an institution must establish a reasonable satisfactory academic progress policy for determining whether an otherwise eligible student is making satisfactory academic progress in his or her educational program and may receive assistance under the title IV, HEA programs". This new regulation was in effect starting July, 1, 2011,

will be driven out the system as shown by the shifting of the demand curve inward from D_0 to D_1.

The net impacts of supply increases and students' loan decreases on college tuition, competition and state subsidy will depend on the directions and magnitude of the elasticity of demand. If demand does not grow fast enough and without the government's assistance, then the tuition will be forced to go down. But, it never happens in the real life, at least for now[4]. So far, the government's assistance has kept the colleges and universities from collapsing. The question that one might ask is, how long and how much resources the government has to support the college subsidy? The 2011 regulation on Title IV can be seen as the product of continued depleting resources which have forced the government to introduce the SAP regulations. The new rules will add the pressure on HEIs to operate more efficiently and to be on their own feet. The regulator has no other choice, but to encourage the use of student's loan more effectively. Without or with less support from the regulator, there are pretty good chances that three things might happen in the long-run i.e., consolidation or merger, college closures or direct investment from other countries. HEIs may not need to face these choices, if the problem facing the US colleges and universities are a short-run problem by taking commercial loans. Recent fundamental and structural changes in the industry will definitely affect the whole system and every single player in the industry many years to come. Therefore, it surely has long-run impacts and consequences on the way the schools' administrators manage their institutions.

[4] However, if this scenario occurs, it is going to be a real painful. It will create larger impacts to the society than the effects of the 2008 financial crises.

This discussion shows that HEIs in the US are facing serious and real challenges. Obviously, there is a need to find new concepts, tools, methods and ways such as the IRI to prevent the HEIs from collapsing. Only the most flexible and adaptive HEIs will survive the recent changes.

Things are getting gloomier when one considers the fact that the increase in supply for education services has accelerated in the fastest rate because of the technological changes. This happens when regulators let for-profit institutions to participate in the industry. Technological changes enable the for-profit and not-for profit HEIs to offer courses and degree online. This new development has caused the supply for education services to increase as shown by the shift of the supply curve further outward/to the right. Past data showed that new institutions that offered both courses and degree online and operate in various locations/campuses have experienced an exponential student enrollments growth, while the traditional players' enrollment growth is below the national average. For example, Daymar College in Ohio has an average enrollment growth (above national growth) of 31% during 2007 to 2011 academic year (Source: Harjanto Djunaidi[5])

Figure 1 shows that as the industry gets more saturated, the subsidy needed to enroll one more student increases as measured by ($T_0 - T_2$). As more institutions enter the same industry and offer the same type of services, then one might expect the risk of going under will also increase. If

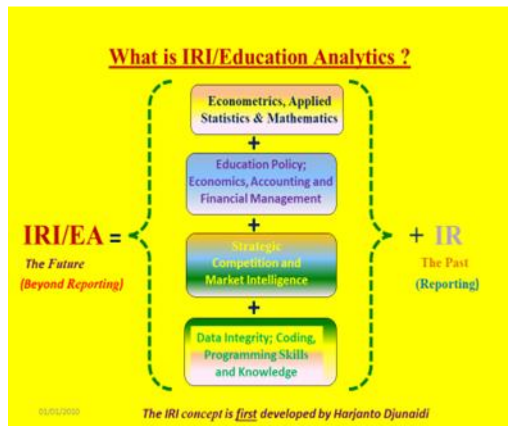
[5] Visit http://www.aaea.us/ for more information on the topic.

demand cannot keep up with the accelerated increase in the supply then the risk of possible college closures will also move toward the same direction with the increase number of the players in the market.

The simple comparative static analysis shows that without the government's help, US colleges and universities need to work extra hard to keep the ship from sinking. There is no other choice for both private and public colleges and universities to escape. Given the new facts, there are strong interests from some of HEIs to find a new and better way to manage their resources. There is a clear evidence that shows the top private universities and colleges are pretty aware of the situation while smaller ones are still operating as usual (BAU—business as usual).

Objective

The purpose of this study is to show how HEIs might be able to manage their resources better by applying IRI[6]. This new concept provides tools or combination of approaches such as statistical analyses, econometrics, mathematical programming, marketing research/intelligence, computer programming and others as shown in Figure 2 to solve HEI's problems. IRI approach provides the policy and decision makers with strategic information which can be used to improve different important education metrics such as retention and graduation rate, students' enrollment, financial aids, early alerts and others. An example of monitoring students' success using order logistic regression will be used in this paper to show one among many possible IRI applications.



Supposed the senior leadership team (SLT) has the interest to know if there is a negative relationship between the number of taken credit hours and students' performance (grade). From the SLT's point of view, a negative correlation will definitely affect the ability of students to graduate. This, in turns will have unfavorable effects on graduation and retention rates, students' financial burden, tax payers, school's reputation and others.

Data

There are 400 observations used in this study. This is a created data set and it

[6] The IRI concept was first developed by *Harjanto Djunaidi* in 2010. The concept was first introduced and presented at 2012 North Carolina Community College System Annual Conference in Raleigh, NC on October 8, 2012 and at 2012 South Central SAS Users Group Annual Conference in Houston, TX on November 5-6, 2012. Applications of IRI to cope and adapt with recent dynamic changes in the industry and to increase the efficiency and effectiveness of students' recruitment, persistent and graduation rate are among many chapters discussed in his upcoming book—*Institutional Research Intelligence: Go beyond Reporting*. The electronic version of the book will be available on the following website: http://www.aaea.us/. Author's email: aaea.us@yahoo.com

does not represent any institution. There are 10 variables originally, but only 3 will be used in the final estimation. These variables are (1). LG = Students' letter grade (2). H = Number of taken credit hours and (3). A = Age.

There are two types of variables in this study—categorical (LG) and continuous (H and A). LG has 8 categories which are A, B, C, D, F, W (withdrawn), WI (instructor withdrawn) and AU (audit). There are many different ways to answer the SLT's question. For example, one can check Pearson correlation (Pearson's r) statistics using PROC CORR in SAS®. However, the r only looks at one-to-one or linear correlation (between two variables) and it will not be able to take into account the interdependency across the variables under studied. Therefore, one needs to apply some sort of regression analysis in order to get the research done right. The SLT has requested to make the data talk so that strategic decisions can be made appropriately. The traditional IR cannot answer the question precisely as requested. Though the question sounds simple, it is more than running a regression analysis and pop with the answer. It requires a deeper understanding of statistical analyses, research methods, marketing research approaches and both modeling expertise and experience. The dependent variable has 8 categories. Therefore, the only appropriate way to test the maintained hypothesis is by estimating a logistic regression.

The Model

The logistic regression is applied to investigate the effects of a set or a group of explanatory variables on a dependent variable (DV). In this case, the DV has eight categories. Therefore, a multinomial logistic model (MLM) is appropriate to be used. However, MLM will produce complex results and difficult to interpret. For this reason, in this paper an order logistic model (OLM) will be estimated instead of MLM. The OLM requires that the DV has a value of either 1 or 0 which is not available in the data set. It needs to be created. In this paper, the target (a grade ***other*** than A, B or C) will be coded as "1". The reference group (a grade equals to A, B or C) is coded as "0". AU grade is excluded in the analysis for two reasons: (1). It does not really belong into either group (2). It has a smaller number of observations. The probability of a student to fail a particular course equals to the odds of the exponential function of the linear regression equation ($\alpha + \beta x$) as shown in equation (1). Equation (2) is known as the Logistic/Logit function. It represents the log odds or natural logarithm of the odds and it serves as a link function between the odds and the linear regression equation. Equation (3) shows the odds that students may fail the course equals to the exponential function of the linear regression equation.

$$(1)\ \pi(x) = \frac{e^{(\alpha+\beta x)}}{1+e^{(\alpha+\beta x)}}$$

$$(2) Ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta x$$

and

$$(3)\ \frac{\pi(x)}{1-\pi(x)} = e^{(\alpha+\beta x)}$$

π (x) is the probability of an event x ("the odds") has a value equals to 1. In this study x represents fail grade category (x=1). The parameter α and β are intercept and regression coefficient, respectively. Ln and e are natural log and exponential function. Knowing each student's probability [π(x=1)] to fail a course in advance gives the Office of Students Success (OSS) or academic counselors/advisors the ability to possibly affect the learning outcome. Therefore, they may not need to spend extra time to monitor their advisees. If the number of successful students increases, it certainly has direct favorable impacts not only on the important metrics such as retention or graduation rate, but also on school's financial situation. The ability to predict the probability that certain group of students will or will not fail their courses is becoming more important for strategic planning purposes especially in the new competitive era.

The most common practice that most schools has been doing in the past is to meet the advisee at the beginning of each semester and left the student by her or himself the rest of the semester. This practice while unfortunate cannot be changed drastically in the future if there is no additional resources are available to hire more advisors. IRI is able to fill the gap and satisfies the urgent need of making more efficient decisions. It helps predicting students' success before the classes' even start. IRI is capable to identify the "potential trouble" students group and send them and their advisors an early alert automatically, before it is too late. At the four-year colleges, the need to monitor the students' academic progress is important, but it is less urgent compare to those at the community colleges for their students are expected to be more prepare for rigorous college level courses. However, the same IRI concept can be applied at four-year colleges as well.

Result and Discussion

Table 2 shows the estimation results. Though the sign is correct, the estimated coefficient shows the number of taken credit hours variable is not significant. This implies the number of taken credit hours may not have a significant effect on students' performance. But, it does show a negative correlation among these two variables. Had more observations were used in the analyses, it may change the estimated results. Age variable is significant at a five percent confident level. The grade variable has an inverse relationship with age. The results suggest that younger students may tend to perform better than their older counterpart. The concordance or C-statistic measures how well the model is able to discriminate between observations at different levels of the outcome. Hosmer and Lemeshow consider c values of 0.7 to 0.8 to show acceptable discrimination. C values of 0.8 to 0.9 indicate excellent discrimination while any values greater or equal to 0.90 shows outstanding discrimination. The C statistics value is 0.585 which is below the threshold for acceptable discrimination. Had the C statistics value falls within the acceptable range, then the model is capable to differentiate the student population into two groups—trouble or safe group.

Table 2 -Logistic Results

Variable	Estimate	Chi_Square	Pr > Chisq	Notes
Intercept	3.5778	3.6308	0.0567	
LH	-0.5855	1.5489	0.2133	Log of taken credit hours
A	-0.1555	4.2096	0.0402	Age
C-Statistics = 0.585				

The "trouble group" of students needs extra help and be monitored closely on their school performance throughout the semester. If things get worse then an early alert can be sent before end of the semester. This example has demonstrated the potential use and the benefits of applying the new *Institutional Research Intelligence* concept instead of the traditional IR.

ROC Graph

The ROC (Receiving Operating Characteristic) is a graph which helps the decision makers to identify the potential benefit of applying a binary model in the decision making process. In this study, it visually shows the effects of implementing the model on the ability to identify students who may fail their classes. The horizontal axis is labeled as (1 - Specificity) and the vertical axis is called Sensitivity. Both Specificity and Sensitivity have a value between 0 and 1.

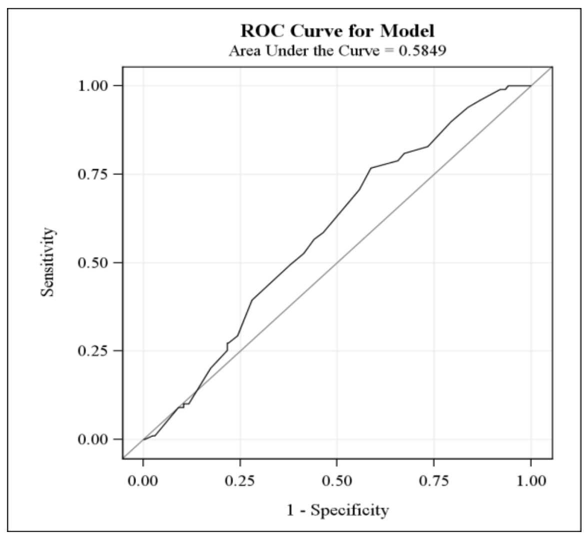
As previously discussed, the dependent variable in the logistic modeling has two possible values. The value equals to "1" if students fail a course (F), otherwise the value equals to "0" if the student pass the course (P). In such a case, there are four possible outcomes. If the model prediction is F and the actual event is also F, then one might call it correct F (CF). If the actual event is P, but the prediction is F then it is said false F (FF). Likewise, if the predicted outcome and actual event matches in the case of P, then it is called correct P (CP). Otherwise it is called false P (FP). One can explain this concept more easily using a 2by2 contingency table as shown in Table 3.

Table 3 - Fail and Pass Course Contigency Table

		Actual Outcome	
		Fail Course (FC)	Pass Course (PC)
Prediction	(FC)'	Correct Fail (CF) = 40	False Fail (FF) = 30
Outcome			
	(PC)'	False Pass (FP) = 20	Correct Pass (CP) = 30
	Total	60	60
CFR = 66.67%			
FFR = 50%			

The ROC graph as shown in Figure 3 can be drawn by plotting the relationships between FFR (False F Rate) and CFR (Correct F Rate) as x and y axes. This graph represents the relative trade-offs between CFR and FFR. The x (1 - Specificity) axis represents FFR and the y (Sensitivity) axis depicts the CFR. The (0, 1) coordinate of the ROC graph shows a perfect correct prediction of students who fail their classes with no FF prediction i.e., perfect model. This point is also known as perfect classification. The diagonal or tangent 45 degree line (which equals to 1) or no-discrimination line represents a random guess of which enrolled students will not pass their course. The area above the diagonal line represents the ability of the model to predict the number of students who will be classified as fail students. Perhaps, the better way to interpret the tangent 45 degree line in this case as the ability of the OSS to predict randomly how many students will fail their courses (***random guess***). The IRI approach adds the ability of the OSS office to forecast more accurately.

Figure 3 – Students Enrollment ROC Graph



IRI application helps the office be able to affect the students' learning outcome at the individual level. The area under the ROC (AUC) graph indicates the probability of the model to correctly rank the pair of (CF, FF). In other words, it is the probability of correct pairwise rankings. This example certainly shows a clear benefit of applying IRI approach in making strategic decision such that the institution might be able to increase the retention and graduation rates. This is true because the model is capable to flag and identify potential troubled students throughout the semester. Early warning enables the decision makers to take extra actions or to apply policies which may potentially affect the final outcomes.

References

Harjanto Djunaidi., *Institutional Research Intelligence: Go Beyond Reporting*. Upcoming Book. Summer 2013.

Harjanto Djunaidi., Predicting Students Enrollment Using SAS. *2012 South Central SAS Users Group Conference*. Houston, Texas, November 5-6, 2012.

Harjanto Djunaidi., Data-driven Decision: Managing Higher Education Institutions more Effectively. *2012 North Carolina Community College System Conference*. Raleigh, October 7-9, 2012.

Harjanto Djunaidi., Institutional Research Intelligence: Go Beyond Reporting. *2012 South Central SAS Users Group Conference*. Houston, Texas, November 5-6, 2012.

Logistic Regression Internet Access: http://en.wikipedia.org/wiki/Logistic_regression

Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York, NY: John Wiley & Sons, Inc; 2000.

National Center for Education Statistics (NCES). Internet Access (05/01/2012).

SAS/STAT® 9.3

34 CFR 668.34-Satisfactory academic progress. Internet access: http://www.ifap.ed.gov/qahome/qaassessments/sap.html

Table 1 - The Number of Higher Education Institutions in the US (2005-2010)

States	Year			Changes in Numbers			Changes in Percent		
	2005	2008	2010	2005-08	2008-10	2005-10	2005-08	2008-10	2005-10
Alaska	12	10	12	-2	2	0	-17%	20%	0%
Alabama	81	81	89	0	8	8	0%	10%	10%
Arkansas	83	84	87	1	3	4	1%	4%	5%
American Samoa	1	1	1	0	0	0	0%	0%	0%
Arizona	114	119	129	5	10	15	4%	8%	13%
California	699	692	747	-7	55	48	-1%	8%	7%
Colorado	110	112	123	2	11	13	2%	10%	12%
Connecticut	102	105	115	3	10	13	3%	10%	13%
District of Columbia	23	23	25	0	2	2	0%	9%	9%
Delaware	17	18	21	1	3	4	6%	17%	24%
Florida	308	332	387	24	55	79	8%	17%	26%
Fed States of Micronesia	4	1	1	-3	0	-3	-75%	0%	-75%
Georgia	178	184	184	6	0	6	3%	0%	3%
Guam	3	3	3	0	0	0	0%	0%	0%
Hawaii	29	28	27	-1	-1	-2	-3%	-4%	-7%
Iowa	93	94	92	1	-2	-1	1%	-2%	-1%
Idaho	28	27	34	-1	7	6	-4%	26%	21%
Illinois	288	282	305	-6	23	17	-2%	8%	6%
Indiana	132	143	156	11	13	24	8%	9%	18%
Kansas	87	89	89	2	0	2	2%	0%	2%
Kentucky	113	109	111	-4	2	-2	-4%	2%	-2%
Louisiana	157	152	155	-5	3	-2	-3%	2%	-1%
Massachusetts	193	188	202	-5	14	9	-3%	7%	5%
Maryland	92	92	98	0	6	6	0%	7%	7%
Maine	41	41	41	0	0	0	0%	0%	0%
Marshal Islands	1	1	1	0	0	0	0%	0%	0%
Michigan	174	179	198	5	19	24	3%	11%	14%
Minnesota	132	137	141	5	4	9	4%	3%	7%
Missouri	213	208	217	-5	9	4	-2%	4%	2%
N. Mariana Islands	1	1	1	0	0	0	0%	0%	0%
Mississippi	63	62	63	-1	1	0	-2%	2%	0%
Montana	34	31	32	-3	1	-2	-9%	3%	-6%
North Carolina	161	167	182	6	15	21	4%	9%	13%
North Dakota	27	29	29	2	0	2	7%	0%	7%
Nebraska	52	53	54	1	1	2	2%	2%	4%
New Hampshire	42	46	44	4	-2	2	10%	-4%	5%
New Jersey	156	154	160	-2	6	4	-1%	4%	3%
New Mexico	50	48	51	-2	3	1	-4%	6%	2%
Nevada	35	36	44	1	8	9	3%	22%	26%
New York	462	462	466	0	4	4	0%	1%	1%
Ohio	339	426	409	87	-17	70	26%	-4%	21%
Oklahoma	138	141	145	3	4	7	2%	3%	5%
Oregon	92	88	88	-4	0	-4	-4%	0%	-4%
Pennsylvania	426	400	403	-26	3	-23	-6%	1%	-5%
Puerto Rico	163	156	155	-7	-1	-8	-4%	-1%	-5%
Palau	1	1	1	0	0	0	0%	0%	0%
Rhode Island	25	23	24	-2	1	-1	-8%	4%	-4%
South Carolina	89	94	104	5	10	15	6%	11%	17%
South Dakota	31	30	31	-1	1	0	-3%	3%	0%
Tennessee	158	172	177	14	5	19	9%	3%	12%
Texas	405	405	442	0	37	37	0%	9%	9%
Utah	59	67	75	8	8	16	14%	12%	27%
Virginia	158	153	166	-5	13	8	-3%	8%	5%
Virgin Island	1	1	1	0	0	0	0%	0%	0%
Vermont	30	29	29	-1	0	-1	-3%	0%	-3%
Washington	124	120	126	-4	6	2	-3%	5%	2%
Wisconsin	93	106	118	13	12	25	14%	11%	27%
West Virginia	83	77	80	-6	3	-3	-7%	4%	-4%
Wyoming	12	13	12	1	-1	0	8%	-8%	0%
Total	9023	9134	9513	108	377	485	1.20%	4.13%	5.10%

Source: National Center for Education Statistics (Web accessed on 05/01/2012)