An Item Selection Strategy Based on Association Rules and Genetic Algorithms

Ming-Hsiung Ying, Shao-Hsuan Huang, Luen-Ruei Wu
Chung Hua University, Taiwan
E-mail: {mhying, M09610023, M09610028}@chu.edu.tw

***Abstract*—The online learning and testing have been as important topics of information education. The main purpose of academic testing is to improve learning. Students could take online test to evaluate their achievements to learning goals. Many online test systems randomly generate test papers from an item bank. A high-quality test paper must to consider the following questions. Is the depth and breadth of test items appropriate? Can test items examine student ability at different cogitative levels? Can test items avoid relationships among test items? Can a test identify student ability and provide learning suggestions appropriate? Therefore, it is the important issue to solve above problems by using information technology. This study applies a novel item selection strategy implemented by computer and is based on assessment theory, association rule, genetic algorithms and a revised Bloom taxonomy. The proposed strategy ensures that test is high quality.**

***Index Terms*—item selection strategy, association rule, genetic algorithms, revision of Bloom's taxonomy, assessment theory**

I. INTRODUCTION

The online learning and testing have been as important topics of information education. Test is a kind of tool to evaluate students' achievement, and it can also obtain large amounts of data related to learning progress and are a fair way to evaluate student learning status because subjective factors do not affect test results. Therefore, testing can enhance learning performance, and it is the best way to evaluate student learning status [4].

With the rapid development of information technology in recent years, there are numerous benefits for computer-based testing (CBT) and online testing (OLT) delivery, including data-rich test results, immediate test feedback, convenient test times and places, reduction in the time-consuming job of scoring tests and preparing items , test items can be created in 'banks' and delivered at random [14][8][13]. Therefore OLT and CBT have become important issues in information education.

In 1956, Bloom identified a learning taxonomy. The Bloom's taxonomy has been used by instructional designers and teachers at all levels of education. Anderson and Krathwohl revised Bloom's original taxonomy by combining both the cognitive process and knowledge dimensions in 2001. This new expanded taxonomy can help instructional designers and teachers to set meaningful learning objectives, and it provided the measurement tool for thinking [3].

In recent years, many researchers applying genetic algorithms (GA) to test paper generation, GA can improve test paper generation operational efficiency [15].

This study uses assessment theory, association rules, GA and a revised Bloom taxonomy to ensure that tests paper is high-quality. To determine whether test items have appropriate difficulty and extensively knowledge content, one must ensure the correlation coefficient between items is minimized, and evaluate different identification knowledge level items.

The proposed item selection strategy considers various factors that can affects item appropriateness and subject relevance of parameters such as difficulty, discrimination, distribution rate of the revised of Bloom taxonomy, and the content of items.

The purposes of this study summarized as follows.

1. Use association rules, assessment theory, genetic algorithm, and revised Bloom taxonomy to ensure that tests paper is high quality.
2. Consider some item-related factors that may affect the test quality, such as difficulty, discrimination, revised Bloom taxonomy distribution rate, content of items, and ensure the correlation coefficient between items is minimized, and evaluate different identification knowledge level items.
3. Achieve efficient management of tests and maintain a bank of items for measuring student learning performance.
4. A preliminary study of the item selection strategy confirmed that the computer automatically generated high-quality test papers.

II. RELATED WORK

The purpose of education is to lead to cause students to change their behaviors, and teachers have to use tests well to assess their change condition. The main purpose of tests is to improve the learning performance [5]. Test theory is primarily divided into classical test theory and modern test theory. Classical test theory is based on the true score model [12][19], whereas modern test theory is based on item response theory[20]. Therefore, via a "self-study" of online learning environments, learners can use online tests to evaluate their performance. Although students can get a test score, the most important factor is to combine the items with educational goals. To achieve this objective, an online testing system is included in the information and knowledge level data for each item.

doi:10.4304/jsw.5.12.1378-1383

Yu argues that item analysis focuses on statistical characteristics with quantitative evaluations, and these characteristics include both discrimination and difficulty [24]. Via analysis teachers can gain information for various items to develop high-quality test papers, indirectly enhancing test reliability and validity, and the selection of test items from an item bank. Teachers can use the test results to implement item analysis and retain excellent items for future reference. To achieve time and energy efficiency, teachers can combine test and evaluation functions to improve their teaching.

Association rules are typically utilized in data mining. Data mining is a technique used to find information in hidden knowledge. For instance, a retail store manager determined that customer will buys soft drinks and snacks together. A strategy will place the two products near each other. Therefore, sales of both products can be increased.

Holland's GA is based on Darwin's evolution theory. The goal is to find optimal solution algorithms [16]. This algorithm is based on the evolutionary concept that animals and plants suited to an environment will outlive those are unsuited to the same environment. Through the designed fitness function, the offsprings will continue to mate and produce offsprings that are relatively more suited to the environment [7].

A GA consists of five components, and these five components are [9]:

1. A method for encoding potential solutions into chromosomes.
2. A means of creating the initial population.
3. An evaluation function that can evaluate the fitness of chromosomes.
4. Genetic operators that can create the next generation population.
5. A way to set up control parameters, e.g., the population size, the probability of applying a genetic operator, and so on.

This computation process generates an initial population, chromosomes, and genetic coding and decoding. To calculate the value of a fitness function, there are two operations in the traditional genetic algorithm that we adopt in this research:

1. Genetic operations: crossover, mutation, and inversion.
2. Evolution operation: selection.

As shown in Figure 1, crossover operator swaps the left potions of two chromosomes. Mutation operator replaces bits on a chromosome with randomly generated bits.

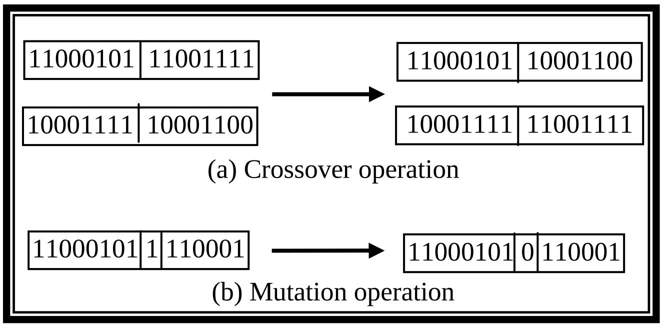

Figure 1. Genetic operators of genetic algorithm

GA are probabilistic search techniques, it has been widely used in engineering and science for solving optimization problems, as well for commercial and financial projections.

Even though the GA based approaches prove to be robust and efficient, another important factor to be considered in making GA efficient is the tuning of control parameter values. According to Grefenstette(1986), the parameters considered in GA include the population size (N), the crossover rate (C), the mutation rate (M), the generation gap (G), the scaling window (W), and the selection strategy (S). In his research, Grefenstette denoted a GA with some specific parameter values as GA (N, C, M, G, W, S)[11].

In designing test items, the educational objectives should be considered. Bloom's Taxonomy divides educational objectives into three "domains:" Affective, Psychomotor, and Cognitive. Within the taxonomy learning at the higher levels is dependent on having attained prerequisite knowledge and skills at lower levels[6][21]. Cognitive psychologists recognize three distinct types of knowledge : declarative, procedural, and conditional. Declarative knowledge is knowing "that". Procedural knowledge is knowing "how" to execute a skill or apply concepts and principles to specific situations. Condiional knowledgr is "knowing when and why" to utilize declarative or procedural knowledge [1][10][22][18].

TABLE I.
The Revised Bloom's Taxonomy

Knowledge Dimensions	Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual						
Conceptual						
Procedural						
Meta-cognitivel						

In recent years, the taxonomy of educational objectives by Bloom et al. has been widely used. Bloom identified six levels within the cognitive domain, including knowledge, comprehension, application, analysis, synthesis and evaluation [3]. Anderson and Krathwohl revised the original taxonomy of Bloom by combining both the cognitive process and knowledge dimensions [3]. Table Ⅰ describes the revised Bloom's taxonomy, the taxonomy comprises a two-dimensional table. One dimension identifies the knowledge (the kind of knowledge to be learned), while the other identifies the cognitive process (the process used to learn). The knowledge dimension comprises four levels: factual, conceptual, procedural, and meta-cognitive. The cognitive process dimension comprises six levels: remember, understand, apply, analyze, evaluate, and create. This new expanded taxonomy can help instructional designers and teachers set meaningful learning objective, and provide the measurement tool for thinking[2][17][23].

III. Research structure

A. Research method

In the marketing field, researchers attempt to find association rules based on the correlation coefficient between commodities and present highly relevant merchandise in the same field to enhance sales. In the educational assessment field, teachers would like the test papers coefficient between items to be as low as possible, items with the appropriate difficulty and a high degree of discrimination. Information technology is combined with test theory, such that an on-line test system can use the selection strategy to implement an online test system.

B. Research steps

The first step in this study is to build an item bank of “Enterprise Resource Planning” course. The item bank contains 313 items that lists four types information of each item: difficulty, discrimination, knowledge level of the revised Bloom's taxonomy, and cognitive process level of the revised Bloom's taxonomy. This experiment adopts three chapters with 69 items because the item bank is too large. The system generates a test paper for students, and according to test results, calculates item difficulty and discrimination. The formulas for item difficulty and item discrimination are as follows:

1. Item difficulty: is derived by DIF=R/N, where the DIF is item difficulty, N is the total number of students who answered this item, and R is the number of students who answered the item correctly.
2. Item discrimination: is derived by $DIS=P_H-P_L$, where DIS is item discrimination. P_H is the percentage of students who had the right answer and in the high score group, and P_L is the percentage of students who had right answer and in the low score group. Test scores are separate into three groups - the high score group, middle score group and low score group; 25-33% of the total number of students are chosen for each group.
3. Correlation coefficient between items: is derived by $R_{a,b}=(CT_{a,b}+ IT_{a,b})/ ST_{a,b}$, where $R_{a,b}$ is the correlation coefficient between item a (I_a) and item b (I_b) , $ST_{a,b}$ is the frequency of I_a and I_b that random chosen into the different student’s test. $CT_{a,b}$ is the frequency of students who answered the I_a and I_b simultaneous correctly, and $IT_{a,b}$ is the frequency of students who answered the I_a and I_b simultaneous incorrectly. $S=\{I_1,I_2,\ldots,I_n\}$ is a combination of item bank, n is the total numbers of items in the item bank, For example, there are 50 students participate in a test, there are 50 different test papers generated. The item 7 and item 9 chosen in a test paper totally 25 times, 12 students answered the I_7 and I_9 simultaneous correctly, and 8 students answered the I_7 and I_9 simultaneous incorrectly The correlation coefficient between item 7 and item 29 is $R_{7,9}$ =(12+8)/25=0.8.

C. Chromosome structure and fitness function design

This study use binary code. Notably, C_j=1 means the item will be used in this test , and C_j=0 means the item will not be used in this test.

Suppose that the item bank has a total of 10 items represented as C_1-C_{10}, if a chromosome structures {C_1, C_2,….., C_{10}} has been written as {101001011000101}, it means that the 2[th], 4[th], 5[th], 9[th] item can be chosen into a test that has 4 items (Figure 2).

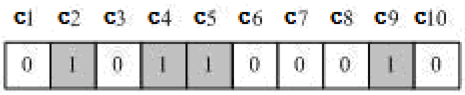

Figure 2. Chromosome structure design

The fitness function design in this study is based on item difficulty and item discrimination of test theory, and a fitness ratio of Bloom’s taxonomy. The following brief outline of the GA illustrates how the GA functions, where notation *S(t)* is the population in the t*th* generation; *si(t)* is the i*th* member in *S(t)*; *f(si(t))* is the fitness value of *si(t)*; and *TOTFIT(t)* is the sum of *f(si(t))* for all *si(t)*. We calculate the quality score (satisfaction degree) using the fitness function described as follows:

$$f(si(t)) = Wdif * ADif(si(t)) + Wdis * ADis(si(t)) + Wac * AC(si(t)) + Wrb * RB(si(t)) \qquad (1)$$

where *ADif* is the *si(t)* chromosome’s functional score for item difficulty, *ADis* is the *si(t)* chromosome’s functional score for item discrimination, *AC* is the correlation between two items, *RB* is fitness ratio of Bloom’s taxonomy, and *Wdi*, *Wdis*, *Wac*, *Wrb* are the weight for each factor. The factors of fitness function are described as follows:

1. Average fitness score of item difficulty-*Adif* and average fitness score of item discrimination-*Adis*: Before calculating the fitness score, *Adif* and *Adis* should be derived; the derivational formula is:

$$ADif\ (s_i(t)) = 1 - \left| \frac{\sum_{j=1}^{x} \left[C_j * PD_j(s_i(t))\right] / y - 0.5}{\sqrt{\sum_{i=1}^{x} \left(PD_i - \overline{PD}\right)^2 / x}} \right| \qquad (2)$$

where *x* is the total number of items in the item bank, and the number of genes in all chromosomes *si(t)*, *y* is the number of items needed in the test paper, $PD_j(s_i(t))$ is the difficulty of the j [th] gene in $s_i(t)$, and C_j=0,1.

$$ADis(s_i(t)) = \frac{\left(\sum_{j=1}^{x}\left[C_j * DI_j(S_i(t))\right]/y + \sqrt{\sum_{i=1}^{x}\left(DI_i - \overline{DI}\right)^2 / x}\right) - \sum_{i=1}^{x} DI_i / x}{2 * \sqrt{\sum_{i=1}^{x}\left(DI_i - \overline{DI}\right)^2 / x}} \quad (3)$$

where $DI_j(s_i(t))$ is the item discrimination of the jth gene in $s_i(t)$, and C_j=0,1.

2. Average correlation coefficient-$AC(s_i(t))$:For each bit in a chromosome where C_j=1, identify the item id, and calculate the correlation coefficient between any two items, the formula as follows:

$$AC(si(t)) = \frac{\left(\sum_{i=1}^{c_2^x} AC_i / c_2^x + \sqrt{\sum_{i=1}^{c_2^x}\left(AC_i - \overline{AC}\right)^2 / c_2^x}\right) - \sum_{j=1}^{c_2^y}\left(AC_j / c_2^y\right)}{2 * \sqrt{\sum_{i=1}^{c_2^x}\left(AC_i - \overline{AC}\right)^2 / c_2^x}} \quad (4)$$

We suppose that C_2, C_4 C_5,and C_9 are a selected combination of one chromosome (Fig. 1), thus, $R_{2,4}$, $R_{2,4}$, $R_{2,4}$, $R_{2,4}$, $R_{2,4}$, are six $R_{a,b}$ C_2^4 =6 kinds combination $R_{a,b}$.

3. Fitness ratio of Bloom's taxonomy-$RB(si(t))$: This study use Bloom's taxonomy theory which has two dimensions. First is knowledge dimension, another is cognitive process dimension. Table Ⅱ describes item distribution for the item bank, those items are classified according to Bloom's taxonomy:

Table II.
Bloom's Taxonomy Distribution For Item Bank

Knowledge dimension / Cognitive Process Dimension		1	2	3	4	5
		Remember	Understand	Apply	Analyze	Evaluate
A	Factual	20	$N_{'A2'}$ =10	10	5	0
B	Conceptual	10	5	8	0	0
C	Procedural	10	0	2	0	0

where x is the total number of items in the item bank. For example, suppose qk='A2' and N_{qk}=10; thus, the probability of 'A2' is P_{qk}=10/80=0.125. To generate a test where x=20, the expected value E_{qk}=20*0.125=2.5;, to round off decimals, E_{qk} =3. E_{qk} =3 represents the hope that the "factual" and "understand" dimensions appear 3 items in the test. However, E_{qk} >0 and E_{qk} <0.5, one must ensure that at least E_{qk}=1. When $\sum E_{qk} > x$, one must consider Max(E_{qk}), meaning that E_{qk}-1 for the number of items qk='A2' until $\sum E_{qk} = x$. The formula for $RB(si(t))$ *is* as follows:

$$RB(s_i(t)) = \left[\sum_{q=1}^{3}\sum_{k=1}^{5}\left(1 - \left|\frac{E_{qk} - T_{qk}}{E_{qk}}\right|\right)\right] / 15$$

where T_{qk} is the number of items selected in $s_i(t)$.

Thus, we expect that fitness value is as large as possible, to ensure that test quality is good and the test can evaluate student learning performance.

IIIV. Implementation of the Genetic Algorithm

The implementation steps for the GA are as follows:

Step 1

Generate the initial population $S(t)$ in Table Ⅲ, where t=0. The GA parameter is GA (30, 0.975, 0.075, 1.0, 1, E), it means that every generation has 30 chromosomes (POPSIZE=30); the crossover rate is CR=0.975, mutation rate is 0.075, and the number of generations is GENER=300. The weight of each factor is *Wdif*=*Wdis*=*Wac*=*Wrb*=25.

Table III.
Initial GA Population

No	$s_i(t)$	$f(s_i(t))$	$P(s_i(t))$	Accumulative Probability
1	0001001101000000110011010100100000000000000100001100000010011001101000	67.275	0.03749	0.03749
2	0000100100001011110101101000000000100100001001000100001000100011100000	63.436	0.03160	0.06910
	:			
8	11101100100000000010101100011100000000000000001101000000011000100000100	42.847	0	0.49857
	:			
26	01000101000000001001100110000010000010111010000010000111000000000010100	73.272	0.04670	0.26414
	:			
30	0100010001011000000011001100000110100101000100010010010000000000011000	69.687	0.04120	1

Step 2

Calculate the fitness value for each member, $f(s_i(t))$. The fitness value of each number (see Table Ⅲ), e.g. the twenty-sixth chromosome has the optimal fitness value , and the satisfaction degree is 73.272 %.

Step 3

Calculate the selection probability for each $s_i(t)$. The selection probability is defined as

$P(s_i(t))=[f(s_i(t))\text{-}Min(f(s_i(0)))]$ / TOTFIT, and

$$\text{TOTFIT}= \sum_{i=1}^{\text{POPSIZE}}[f(s_i(t))\text{-}Min(f(s_i(0)))]$$

The 26th chromosome has the maximum fitness value, and has the maximum probability to be selected as a parent; the probability value is 0.04670 (Table Ⅲ). The 8th chromosome has the minimum fitness value, and its selection probability is 0.

Step 4

Select a pair of members (parents) that compare the accumulation probability with a random number (ranging from 0 to 1), and reproduce the chromosome into the new population.

Step 5

Apply the genetic operators (crossover, mutation, and inversion) to parents. Replace the parents with the resulting offspring to form a new population, $S(t+1)$, for the generation t+1. If the size of the new population is equal to 30, then go to step 6, else go to step 4. In our case, every generation including 30 chromosomes, we use the higher crossover rate of 0.975 that can generate more newly structure, and prevent dropping into the local optimal solution. Our mutation rate is 0.075. If the result falls in the optimal local area, we will consider it in the computation of the global optimal area to be a possible final number.

Step 6

If the current generation, $t + 1$= 300, then stop; otherwise, go to step 2.

Finally, we got the optimal fitness value is 82.0916 (Table IV). Its chromosomal structure was {0000010100 00000100110101000011001000100000000010100101100 0100000001110}. The 6th, 8th, 16th, 19th, 20th, 22th, 24th, 29th, 30th, 33th, 37th, 46th, 48th, 51th, 53th, 54th, 58th, 66th, 67th, 68th test items were selected into the test paper. It means that this combination of items has the appropriate difficulty and knowledge levels, item discrimination and a variety of cognitive levels.

Figure 3 shows the stable maximum fitness value after 177 generations. The average fitness value increased rapidly after the initial stage reached a stably state. The minimal fitness descended gracefully with the chromosome mutation. The mutant chromosome had the minimal fitness value. Thus, we believe a high-quality test paper was generated.

Table IV.
Variation Of Fitness Function Value

Generation	Fitness function value
1	73.2723
2	73.5823
3	73.7523
4~18	78.1932
19~49	80.4305
50~58	80.6444
59~165	81.6135
166~176	81.9647
177~300	82.0916

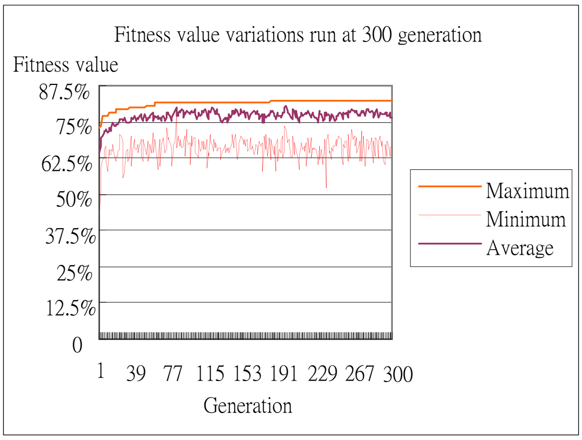

Figure 3. Fitness value of each generation

V. Conclusions

This study applies a novel item selection strategy executed by a computer. The strategy is based on assessment theory, association rule, GA and revised Bloom taxonomy, thereby ensuring that a high-quality test paper is generated.

Through the GA, the item selection strategy generates test papers that meet the needs of most teachers. Additionally, the strategy can easily deal with large amounts of data in the item bank. Furthermore, the item combinations generated by the item selection strategy ensure that the test has appropriate difficulty, high degree of discrimination, a maximum independence parameter between items and an appropriate knowledge level.

Acknowledgment

This research is supported by National Science Council of Taiwan, R.O.C. (Grant No: NSC 98-2410-H-216-006-MY2) and Chung-Hua University (Grant No: CHU- NSC 98-2410-H-216-006-MY2).

References

[1] Anderson, J.R., *The Architecture of Cognition*, Harvard University Press, Cambridge MA ,1983.

[2] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., and Wittrock, M. C., *A taxonomy for learning, teaching,*

and assessing: a revision of Bloom's taxonomy of educational objectives, New York: Longman, 2001.
[3] Bloom, B. S., *Taxonomy of educational objectives: The classification of educational goals, Handbook I: Cognitive domain,* New York:: Longman, 1956.
[4] Brueckner, L. J. & Band, G. L., *The diagnosis and treatment of learning difficulties*, New York: John-Wiley, 1955.
[5] Chang, Y.H. *The test preparation and application*, Taipei : Wei-Wen, 1982
[6] Chang, L.C., *Educational Testing and Measurement*, Taipei : Wu-Nan,1997.
[7] CHEN Y P. *Extending the Scalability of Linkage Learning Genetic Algorithms: theory and practice*, Champaign: UIUC, 2004.
[8] Chou, W.J. "Implementation of computer-assisted testing on WWW," *Proceedings of the Seventh International Conference on Computer Assisted Instruction*, pp. 543-550, 1998.
[9] Davis, L., "Genetic Algorithms and Simulated Annealing. Research Notes in Artificial Intelligence,", Los Altos, CA: Morgan Kaufmann, 1987.
[10] Garavalia, L. S., Marken, P. A., Sommi, R. W.,"Selecting Appropriate Assessment Methods: Asking the Right Questions," *American Journal of Pharmaceutical Education* , Vol. 66, pp.108-112, Summer 2003.
[11] Grefenstette, J. J., "Optimization of Control Parameters for Genetic Algorithms", *IEEE Trans. Systems Man Cybernet.*, SMC-16, pp. 122-128, 1986.
[12] Gulliksen, H., *Theory of mental test*, NJ: Lawrence Erlbaum Associates, 1987.
[13] HE J. K., "The test design of computer learning system", *Secondary Educatio*n，Vol.41(2), pp. 29-34, 1991.
[14] Ho, R.G., "Testing design in computer instruction system," *Secondary Education*, vol.41, no.2, pp. 29-34, 1990.
[15] Huang, M., Tang, L., Liang, X., "Research on Intelligent Test Paper Generation Based on Improved Genetic Algorithm," *2009 Chinese Control and Decision Conference (CCDC 2009)* . China, pp.2967-2970, 2009.
[16] Holland, J. H., "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology," *Control, and Artificial Intelligence,* University of Michigan Press, 1975.
[17] Krathwohl, D. R., "A revision of Bloom's taxonomy: An overview, Theory Into Practice", Vol.41(4), pp. 212-219, 2002.
[18] Kreber, C., "Learning Experientially through Case Studies: A Conceptual Analysis," *Teaching in Higher Education*, vol.6(.2), pp.217-28, 2001.
[19] Lord, F. M. and Novick, M. R., *Statistical theories of mental test scores, Reading*, MA: Addison-Wesley, , 1968.
[20] Lord, F. M., *Applications of item response theory to practical testing problems,* Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.
[21] Orlich, C., Harder, R., Callahan, R., Trevisian, M., & Brown, A., *Teaching strategies: A guide to effective instruction, 7th ed.,* Boston: Houghton Mifflin Company, 2004.
[22] Woolfolk, A.E., *Educational Psychology*, 7th ed., , Boston:MA, Allyn and Bacon, 1998.
[23] Yeh L. C. and LIN S. P., "The research of Bloom's taxonomy for educational objects," *Journal of Education Research* , Vol.105, pp. 94-106, 2003.
[24] Yu M. N., *Testing and assessment in education*, Taiwan: psychology, 1997.

Ming-Hsiung Ying: Dr. Ying is an assistant professor of information management at the Chung Hua University, HsinChu, Taiwan. He received the PhD in Management Information System from National Chengchi University, Taipei, Taiwan in 2006. His research interests center on E-learning, knowledge management, psychology of Internet, database system, and enterprise resource planning.

Shao-Hsuan Huang: He is a Master student at the Department of Information Management, Chung Hua University, Taiwan. His major is information management and E-learning.

Luen-Ruei Wu: He is a Master student at the Department of Information Management, Chung Hua University, Taiwan. His major is information management and E-learning.