Applying IRT to Estimate Learning Ability and K-means Clustering in Web based Learning

Wen-Chih Chang
Department of Information Management Chung Hua University, Hsin-Chu, Taiwan
earnest@chu.edu.tw

Hsuan-Che Yang
Department of Education Technology, Tamkang University, Taipei, Taiwan
arthur@mail.mine.tku.edu.tw

Abstract — **E-learning provides a convenient and efficient way for learning and enriching people's lives. But there is no appropriate way to estimate and diagnose people when they learning with e-learning environment/system. For learning ability estimation issue, Item Response Theory which plays an important role in modern mental test theory is applied. Besides, K-means clustering method is also applied to cluster learner's ability for remedying courses or enhancing courses. We integrate these two theories and propose a combination methodology to solve the estimation and diagnostic issues in e-learning environment. A web-based assist system is provided as well. Experimental data is collected with forty sophomore students studying "Business Data Communication" class at Dept. of Information Management in Chung Hua University in Taiwan. We illustrated the method to observe and estimate the variation of learner's ability. This methodology and system could make some valuable contribution in e-learning related study and society.**

***Keywords* — E-learning, K-means cluster, Item Response Theory, Learner Ability, Assessment**

I. INTRODUCTION

E-learning provides people a convenient and efficient way for learning and enriching their lives. Web based learning is one of the popular way for people to access learning material at any time, any place. Web based learning supports teachers teaching and students learning easily with web pages. However web based learning did not provide well method to assess learners. In traditional education, the teacher can change his/her lecturing style or content flexibly to maximize the teaching quality with students' response face to face. On the contrary, teachers recorded and prepared their teaching materials before classes begin and these materials will be published on the Internet. Teachers have less time to predict students' learning ability thoroughly before web classes begin. But, it is hard to modify learning content or style immediately and flexibly in web based learning environment. Because the traditional teaching behavior is the synchronized occurrence, but in web based learning environment, it is carries on under the asynchronous condition.

Estimate learners' ability is a significant issue in web teaching. How to assist instructor estimate learner's ability and analyze learning records accurately, which provides precious information to adjust the learning contents or learning sequencing more appropriately. Assessment measures and analyzes student performance and learning skill. It also replies feedback to the teacher and student which documents growth or provides directives to improve future performance, is significant to learning and development. Formative assessment plays the role to guide student instruction and learning, diagnose skill or knowledge gaps, measure progress and evaluate instruction. In daily use, teachers apply formative assessment to determine what concepts require more teaching and what teaching techniques or strategies require modification. After a period of learning days, teachers use the result to evaluate instruction strategies and curriculum. Teachers can make some adjustments for better student performance. Assessment focuses on the gap between students performance and instruction goal. Formative assessment which is beneficial to apply on web based learning to gather the learning information could adapt the teaching or the learning to meet the needs of the learner.

The related work section describes Item Response Theory and K-means clustering method we use in our methodology and system, some related studies are also discussed as well. After related work, the main methodology is presented, it delineates how IRT and K-means diagnose students, estimate their learning abilities and make learners in clusters. An assist system is discussed in web based assist system section. Experiment discussion section records the results of our study. Finally, a brief conclusion and future work is drawn.

II. Related Works

With the rapidly development of e-learning, e-learning assessment has become more and more important. Comparing the traditional learning, international e-learning specification, such as SCORM (Sharable Content Object Reference Model) [1], IEEE LOM (Learning Object Metadata) [2] assisted teachers to standardize their teaching materials. Some assessment methods and systems are proposed. Chang's study of web based assessment [3] is one example of research studies which fall into the field of assessment for distance learning. In brief, some researchers concentrate on developing communication tools and group cooperation while others concentrate on analyzing and evaluating student's learning performance. A courseware design tool with the theory of concept and influence diagram coupled with a user-friendly interface is proposed [4]. The transformation algorithm is also included for the conformance with e-learning standards. With the proposed mechanism and tools, the advantages of courseware diagram are preserved. Students' learning performance can be improved by taking different levels of remedial courses based on student performance with a systematically built course sequence. Some integrated standard and assessment method tool is created. The developed tool can be used for SCORM assessment in three perspectives: student-problem, course-problem, and student problem [5]. So any test problem set and the student performance can be thoroughly examined. The tool was applied to a class and the empirical results are presented in this paper.

Item Response Theory (IRT) is often referred to as latent trait theory, strong true score theory, or modern mental test theory and is distinguished from Classical test theory [6]. Theoretically, IRT is based on two concepts.

(1) The possibility of one student who answered the individual question can be predict or explain by one set of factors.

(2) The relation of the possibility of one student who answered the individual question and the set of factors can be explained by a continuous increasing equation called item characteristic curve.

The definition of item characteristic curve is the possibility when the student answered the question correct. When the item characteristic curve is high means the possibility when the student answered the question correct high. In IRT theorem, each question has only one item characteristic curve which is composed of one or more parameters to describe the question's characteristics. Therefore, item characteristic curve will be different when we apply different parameter equations. For instance, we pick three items (problems) and one-parameter model for the example. In the Figure 1, it represents three item characteristic curves of same discrimination of 1 and distinct difficulties of 1, 3 and 5. In the Figure 2, it also represents three item characteristic curves with same difficulty of 3 and distinct discriminations of 0.3, 0.5 and 0.7.

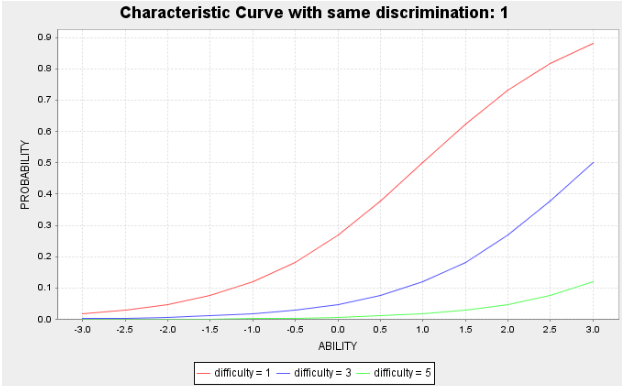


Figure 1 Item characteristic curves with one-parameter model. Same discrimination of 1 with distinct difficulties.

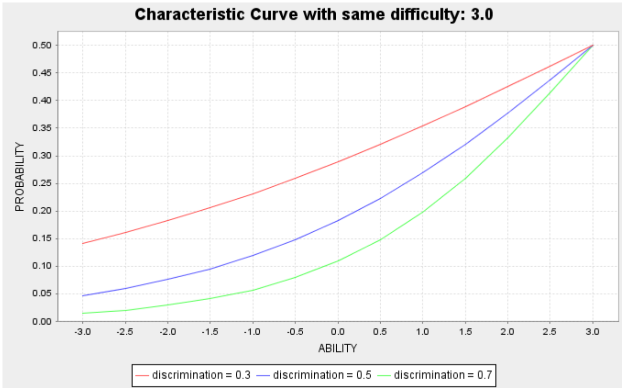


Figure 2 Item characteristic curves with one-parameter model. Same difficulty of 3 with distinct discriminations

Logistic function is the formula to represent characteristic curve of each item in IRT. We can show the logistic function as follow [6].

$$P(\theta) = \frac{1}{1+e^{-a(\theta-b)}}$$

where:
e is the constant 2.718.
b is the difficulty parameter, typical value have -3≦b≦3.
a is the discrimination parameter, typical value have -2.8≦a≦2.8.
L = a(θ- b) is the logistic deviate (logic) and θ is an ability level.

However, there are some special models in basic logistic function, including:

(1). One-parameter logistic model:

One-Parameter logistic model ignores the discrimination of item, and set the discrimination of each item to 1. In this model we can transform the basic logistic function into the formula below [6].

$$P(\theta) = \frac{1}{1+e^{-a(\theta-b)}} = \frac{1}{1+e^{-1(\theta-b)}}$$

(2). Two-parameter logistic model:

Two-Parameter logistic function considers both the discrimination and difficulty of item. And its logistic function is the original type of logistic function.

(3). Three-parameter logistic model:

Besides the discrimination and difficulty of item, Three-Parameter logistic function also considers the guessing factor when people answer the item. The parameter c is the probability of getting the item correct by guessing. After we add the guessing factor into the original logistic function, we get the tree-parameter logistic function as shown [6].

$$P(\theta) = c + (1-c)\frac{1}{1+e^{-a(\theta-b)}}$$

The parameter c has a theoretical range of 0≦c≦1.0, but in practice, values above 0.35 are not considered acceptable, hence the range 0≦c≦0.35 are used here.

IRT always comes with Computerized Adaptive Testing (CAT). Adaptive testing is used in computer administrated tests to dynamically estimate the examinee level, such as Graduate Record Examination (GRE) and the Test of English as a Foreign Language (TOEFL) [7]. However, this kind of technique is appropriate for testing, not for detecting problem's quality. IRT works when there is a need to determine a student's level of knowledge, but not measuring the student's knowledge in every concept or level in the course. This paper proposes a combinative methodology. With K-means clustering, it will cluster learners with learning ability for remedial course. It will also estimate students' learning abilities with IRT.

Clustering is a method finding an organization in a collection of unlabeled data. It is a process putting items into groups that is similar in some aspects. Those clustering algorithms and applications applied on Biology, Marketing, Earthquake studies and City-Planning and so on. K-means method is one of the famous and common clustering methods.

The k-means clustering was invented by H. Steinhaus in 1956 [8]. K-means methodology is commonly used in clustering techniques. K-means analysis let the user begins with a collection of data samples and attempts to group them into k number of clusters based on certain specific distance measurement. Distance is usually based ob the data attribute, like the price of the product, the score of the student and the time interval and the location of the earthquake.

The important steps involved in the K-Means clustering algorithm are listed below.

(1) K-Means algorithm is started by setting 'k' different clusters.
(2) The distance measurement between each node, within a given cluster, to their respective cluster centroid is calculated.
(3) After the distance between each node to centroid is calculated. Each node can be grouped to each cluster depends on the shortest distance from each node to the cluster centroid.

III. DATA ANALYSIS METHODOLOGY

In this section, we illustrated our method to estimates the learner's ability with forty sophomore students studying "Business Data Communication" class in Taiwan. Estimating learner's ability is never easy. In most cases, teachers are used to estimate learner ability with the score. However, the questions which have different difficulty and discrimination in the exam sheet, students have the same score but might have different ability. In order to estimate learner ability, we apply item response theory to estimate each item. We apply two-parameter logistic model to get the P(θ) which is the probability that an examinee with their ability will give a correct answer to the item. The P(θ) range is 0≦P(θ)≦1. In order to get the item difficulty index and item discrimination index, we choose Kelly's method. In 1939, Prof. Kelly indicated that the best percentage is 27%, and the acceptable percentage is 25-33% [9]. We tried to define the percentage 25% in this paper and our system. Following is how we compute item discrimination index in steps.

Step 1: Sort the students order according to the students' score in the exam.
Step 2: P_H is defined as the higher 25% of total students and P_L is defined as the lower 25% of total students.
Step 3: Count a student's correct answers and his/her percentage in the higher group and the lower group of each question.
Step 4: Calculate the item difficulty index for each problem $P = (P_H+P_L)/2$.
Step 5: Calculate the item discrimination index for each problem $D = P_H-P_L$.
Step 6: The following shows the information format we record.

In our experiment, there are twenty items in our exam. The discrimination index and difficulty are listed in Table 1. For example, student No.B09510028. We knew the discrimination index a＝0.4545 and difficulty index b＝0.7727 of item 1 in Table 1. Default learner ability θ is set 1. The calculation progress and the result of student No.B09510028 is shown from Table 2 to Table 5.

Table 1 the discrimination index and difficulty gathered in our exam.

Item No.	Discrimination (a)	Difficulty (b)	Item No.	Discrimination (a)	Difficulty (b)
1	*0.4545*	*0.7727*	11	*0.3636*	*0.7273*
2	*0.4545*	*0.6818*	12	*0.6364*	*0.6818*
3	*0.2727*	*0.8636*	13	*0.3636*	*0.8182*
4	*0.2727*	*0.8636*	14	*0.3636*	*0.8182*
5	*0.4545*	*0.7727*	15	*0.6364*	*0.6818*
6	*0.3636*	*0.8182*	16	*0.2727*	*0.8636*
7	*0.1818*	*0.9091*	17	*0.2727*	*0.8636*
8	*0.2727*	*0.8636*	18	*0.5455*	*0.7273*
9	*0.2727*	*0.8636*	19	*0.1818*	*0.9091*

We put the parameters into the equation and get the P(θ) of item 1. In Table 2, we calculated 20 items for their P(θ).

$$P(\theta) = \frac{1}{1+e^{-a(\theta-b)}} = \frac{1}{1+2.718^{-0.4545*(1-0.7727)}} = 0.525804$$

In the next step, learner's ability is estimated. The learner ability estimation equation is given as following.

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)}$$

$\hat{\theta}_s$ is the estimated ability of the examinee with iteration s.

a_i is the discrimination parameter of item i, i=1,2,….N.

u_i is the response made by the examinee to item i:

u_i = 1 for a correct response

u_i = 0 for a wrong response

$P_i(\hat{\theta}_s)$ is the probability of correct response to item i, under the given item characteristic curve model, at ability level $\hat{\theta}_s$ within iteration s.

$Q_i(\hat{\theta}_s) = 1 - P_i(\hat{\theta}_s)$ is the probability of incorrect response to item i, under the given item characteristic curve model, at ability level $\hat{\theta}_s$ within iteration s.

We count Q(θ), a(u-P) and a*a(PQ) for item 1 below.

$Q_i(\hat{\theta}_s) = 1 - P_i(\hat{\theta}_s)$ = 1－0.525804＝0.474196

a(u-P)＝0.4545×(1-0.525804)＝0.215543863

a*a(PQ)＝0.454×0.454×(0.525804×0.474196)

＝0.51515

With the same calculating progress, we got the result of 20 items for iteration 1 in

Table 2 ability estimation of student No.B09510028 iteration 1. For the reason Δθ s≦0.001 is not equal to our limitation, the iteration would continue.

The first iteration:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)} = 3.707857591$$

The second iteration:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)} = 4.294300844$$

The third iteration:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)} = 4.342182747$$

And the forth iteration:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)} = 4.342182747$$

We set Δθ s≦0.001 because the value of the adjustment is very small. Besides, we checked the standard error which is not greater than 1. The standard error is affected by the quantity of the test item. The more test items, the smaller standard error. We estimate 40 students' learning abilities in Table 6 Experiment result of the learner ability, there are nine students' abilities cannot be estimated. These students who answer all the items correct or wrong, are special cases in our equations. The learner ability is not an infinite sequence toward some limit; the Δθ s will not converge smaller than 0.001. The standard error of Student B09510028 is 0.643251.

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^{N} a_i^2 P(\hat{\theta})Q(\hat{\theta})}} = \frac{1}{\sqrt{0.413772}} = 0.643251$$

Table 2 ability estimation of student No.B09510028 iteration 1

Item	u	P	Q	a(u-P)	a*a(PQ)	$\Delta\theta_s$	θ_s+1
1	1	0.525	0.474	0.2155	0.051515	2.7078	3.7078
2	0	0.536	0.463	-0.2436	0.051384		
3	1	0.509	0.490	0.1338	0.018589		
4	1	0.509	0.490	0.1338	0.018589		
5	0	0.525	0.474	-0.2390	0.051515		
6	1	0.516	0.483	0.1758	0.033022		
7	1	0.504	0.495	0.0901	0.008264		
8	1	0.509	0.490	0.1338	0.018589		
9	1	0.509	0.490	0.1338	0.018589		
10	1	0.509	0.490	0.1338	0.018589		
11	1	0.524	0.475	0.1728	0.032977		
12	1	0.550	0.449	0.2860	0.100209		
13	1	0.516	0.483	0.1758	0.033022		
14	1	0.516	0.483	0.1758	0.033022		
15	1	0.550	0.449	0.2860	0.100209		
16	1	0.509	0.490	0.1338	0.018589		
17	1	0.509	0.490	0.1338	0.018589		
18	0	0.537	0.462	-0.2929	0.07397		
19	1	0.504	0.495	0.0901	0.008264		
20	1	0.516	0.483	0.1758	0.033022		
			SUM	2.0052	0.740514		

Table 3 ability estimation of student No.B09510028 iteration 2

item	u	P	Q	a(u-P)	a*a(PQ)	Δϑs	ϑs+1
1	1	0.791	0.208	0.0947	0.034093	0.5864	4.2943
2	0	0.798	0.201	-0.3628	0.033273		
3	1	0.684	0.315	0.0859	0.016056		
4	1	0.684	0.315	0.0859	0.016056		
5	0	0.791	0.208	-0.3597	0.034093		
6	1	0.740	0.259	0.0942	0.025382		
7	1	0.624	0.375	0.0682	0.007752		
8	1	0.684	0.315	0.0859	0.016056		
9	1	0.684	0.315	0.0859	0.016056		
10	1	0.684	0.315	0.0859	0.016056		
11	1	0.747	0.252	0.0919	0.024976		
12	1	0.872	0.127	0.0809	0.044968		
13	1	0.740	0.259	0.0942	0.025382		
14	1	0.740	0.259	0.0942	0.025382		
15	1	0.872	0.127	0.0809	0.044968		
16	1	0.684	0.315	0.0859	0.016056		
17	1	0.684	0.315	0.0859	0.016056		
18	0	0.835	0.164	-0.4557	0.040873		
19	1	0.624	0.375	0.0682	0.007752		
20	1	0.740	0.259	0.0942	0.025382		
			SUM	0.2854	0.486671		

Table 4 ability estimation of student No.B09510028 iteration 3

item	u	P	Q	a(u-P)	a*a(PQ)	Δθs	θs+1
1	1	0.832	0.167	0.0763	0.028863	0.0478	4.3421
2	0	0.837	0.162	-0.3808	0.028075		
3	1	0.718	0.281	0.0768	0.015053		
4	1	0.718	0.281	0.0768	0.015053		
5	0	0.832	0.167	-0.3782	0.028863		
6	1	0.779	0.220	0.0801	0.022712		
7	1	0.649	0.350	0.0637	0.007529		
8	1	0.718	0.281	0.0768	0.015053		
9	1	0.718	0.281	0.0768	0.015053		
10	1	0.718	0.281	0.0768	0.015053		
11	1	0.785	0.214	0.0780	0.022291		
12	1	0.908	0.091	0.0580	0.03357		
13	1	0.779	0.220	0.0801	0.022712		
14	1	0.779	0.220	0.0801	0.022712		
15	1	0.908	0.091	0.0580	0.03357		
16	1	0.718	0.281	0.0768	0.015053		
17	1	0.718	0.281	0.0768	0.015053		
18	0	0.874	0.125	-0.4772	0.032548		
19	1	0.649	0.350	0.0637	0.007529		
20	1	0.779	0.220	0.0801	0.022712		
			SUM	0.0200	0.419055		

Table 5 ability estimation of student No.B09510028 iteration 4

item	u	P	Q	a(u-P)	a*a(PQ)	Δθs	θs+1
1	1	0.835	0.164	0.0749	0.028447	0.0003	4.3424
2	0	0.840	0.159	-0.3821	0.027663		
3	1	0.720	0.279	0.0761	0.014967		
4	1	0.720	0.279	0.0761	0.014967		
5	0	0.835	0.164	-0.3795	0.028447		
6	1	0.782	0.217	0.0790	0.02249		
7	1	0.651	0.348	0.0634	0.007509		
8	1	0.720	0.279	0.0761	0.014967		
9	1	0.720	0.279	0.0761	0.014967		
10	1	0.720	0.279	0.0761	0.014967		
11	1	0.788	0.211	0.0769	0.02207		
12	1	0.911	0.088	0.0564	0.032742		
13	1	0.782	0.217	0.0790	0.02249		
14	1	0.782	0.217	0.0790	0.02249		
15	1	0.911	0.088	0.0564	0.032742		
16	1	0.720	0.279	0.0761	0.014967		
17	1	0.720	0.279	0.0761	0.014967		
18	0	0.877	0.122	-0.4788	0.031914		
19	1	0.651	0.348	0.0634	0.007509		
20	1	0.782	0.217	0.0790	0.02249		
			SUM	0.0001	0.413772		
					0.643251		

Table 6 Experiment result of the learner ability

Student No.	*Δϑs*	*ϑs+1*	*Standard Error 1/((a*a*(P*Q))^1/2)*	*Iterations*
B09310005	0.000564086	7.228475628	0.425637762	5
B09510002	0.000464682	10.26309262	0.272959423	6
B09510004	TheΔϑs will never converge smaller than 0.001.			
B09510006	0.000157018	4.128891681	0.661535659	4
B09510010	5.98903E-05	6.349052791	0.484351245	5
B09510012	8.18448E-06	3.374039583	0.725815834	4
B09510016	TheΔϑs will never converge smaller than 0.001.			
B09510018	4.23338E-06	8.418639491	0.357142425	6
b09510022	7.85567E-05	3.92651486	0.678904942	3
B09510026	4.41257E-07	7.771627679	0.392841793	6
B09510028	0.000306072	4.342488819	0.643251499	4
B09510030	4.98357E-06	11.76500885	0.220233456	7
B09510038	0.000464682	10.26309262	0.272959423	6
B09510040	4.41257E-07	7.771627679	0.392841793	6
B09510044	1.64374E-07	4.810209551	0.60369427	5
B09510046	5.98903E-05	6.349052791	0.484351245	5
B09510050	0.000564086	7.228475628	0.425637762	5
B09510052	-4.18338E-08	0.394703485	0.859689607	3
B09510054	TheΔϑs will never converge smaller than 0.001.			
B09510056	1.92367E-05	5.981935115	0.510988809	5
B09510060	5.64201E-07	5.06871674	0.582347424	5
B09510062	-2.61989E-08	0.517461764	0.86129777	3
B09510064	7.85567E-05	3.92651486	0.678904942	4
B09510066	0.000184305	6.760413351	0.455989303	5
B09510074	TheΔϑs will never converge smaller than 0.001.			
B09510076	4.23338E-06	8.418639491	0.357142425	6
B09510078	3.82082E-05	3.733991289	0.695390875	4
B09510084	8.18448E-06	3.374039583	0.725815834	4
B09510088	TheΔϑs will never converge smaller than 0.001.			
B09510090	TheΔϑs will never converge smaller than 0.001.			
B09510094	0.000157018	4.128891681	0.661535659	4
B09510098	TheΔϑs will never converge smaller than 0.001.			
B09510102	3.6612E-05	1.890880609	0.830933176	3
B09510104	4.25078E-05	9.218247684	0.317654426	6
B09510106	TheΔϑs will never converge smaller than 0.001.			
B09510108	5.64201E-07	5.06871674	0.582347424	5
B09510202	1.8763E-06	5.347459033	0.559848971	5
B09510204	0.000184305	6.760413351	0.455989303	5
B09510206	0.000564086	7.228475628	0.425637762	5
B09510208	TheΔϑs will never converge smaller than 0.001.			

IV. Web based Assist System

According to the methodology has discussed in previous section. We propose a prototype system to help us to analysis collected data. We use PHP 4.4.5 and Dojo Toolkit 1.0.2 [10] for the JavaScript Library. Apache Web Server 2.0.59 for Http server, MySQL 5.0.27 for the database management system. We build a scorebook system to help teachers to keep scores when having a quiz or examination in semesters. In addition, it also assists teachers in analyzing students' learning conditions and estimating their learning abilities with IRT. This system really reduces the computing time of normal paper work before.

Figure 3 is the snap shot of our proposed system. It can be divided into three parts,

1) Class Selection (upper part of screen)

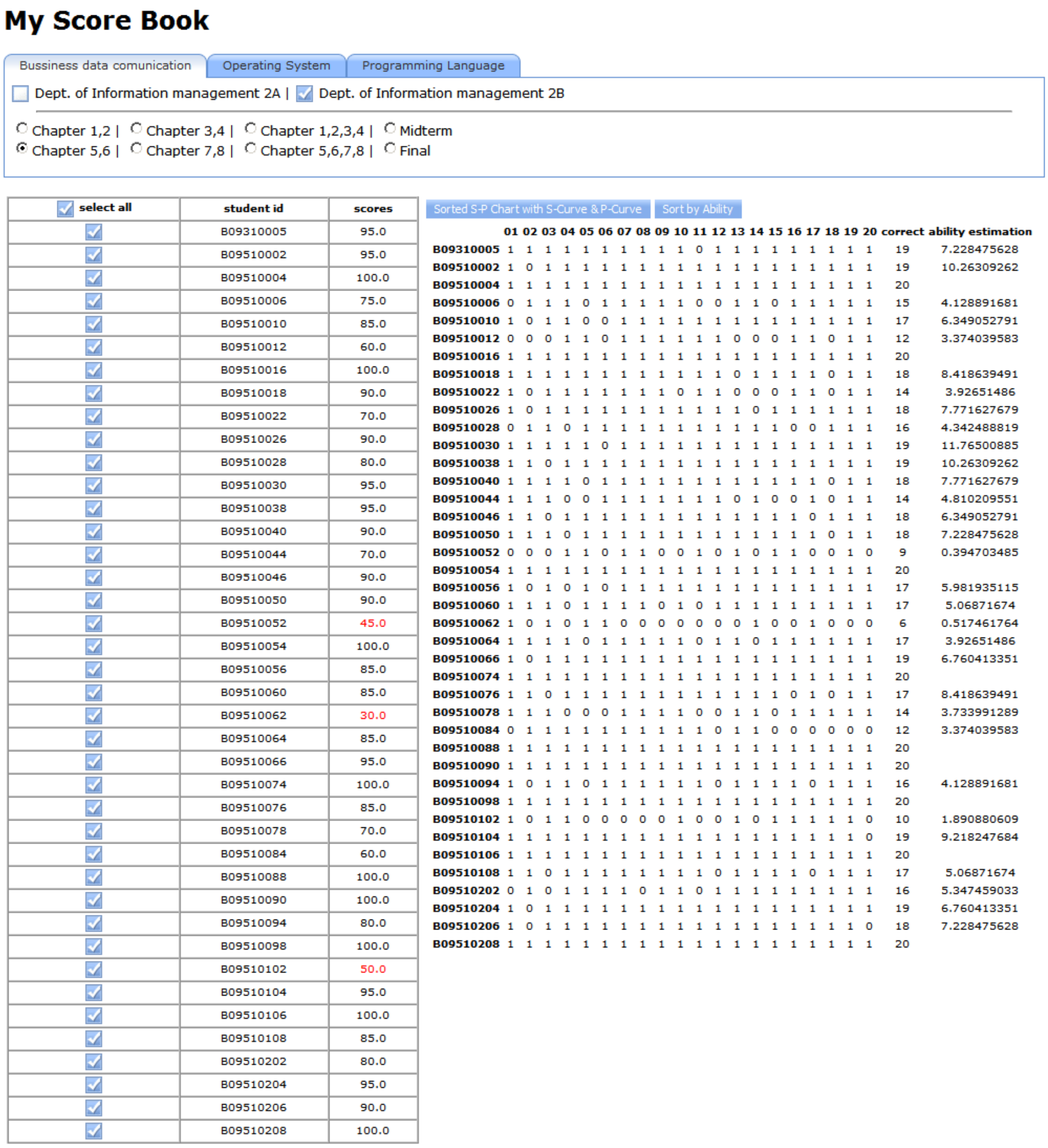
2) Student List and Score book (lower left part of screen)

3) Analyzer (lower right part of screen)

"Class Selection" is displayed when user logon system. After user selects the class, the "Student List and Score book" and "Analyzer" are appearing. Class Selection area shows all curriculums the user teaches or data recorded in tabbed button on top of the screen.

In Figure 4, there are four chapters show on the page, including "Business data communication", "Operating system" and "Programming language". In "Business data communication", there are two classes take this program. And several quiz and examination were held before. When user choose the class and pick the assessment he/she want to analyze, all students attend that assessment will show on the lower left part.

As we can see in Figure 3, All users' id and scores can be listed in "Student List and Score book" as common scorebook system. User can select examinees by selecting checkbox in front of each student. Detail records of examination of selected students will be displayed in the " Analyzer". As we discussed, there are several abilities cannot be estimated and filled with blank. That is because these students answer all the items correct, and it's a special cases in our experiment.



select all	student id	scores
☑	B09310005	95.0
☑	B09510002	95.0
☑	B09510004	100.0
☑	B09510006	75.0
☑	B09510010	85.0
☑	B09510012	60.0
☑	B09510016	100.0
☑	B09510018	90.0
☑	B09510022	70.0
☑	B09510026	90.0
☑	B09510028	80.0
☑	B09510030	95.0
☑	B09510038	95.0
☑	B09510040	90.0
☑	B09510044	70.0
☑	B09510046	90.0
☑	B09510050	90.0
☑	B09510052	45.0
☑	B09510054	100.0
☑	B09510056	85.0
☑	B09510060	85.0
☑	B09510062	30.0
☑	B09510064	85.0
☑	B09510066	95.0
☑	B09510074	100.0
☑	B09510076	85.0
☑	B09510078	70.0
☑	B09510084	60.0
☑	B09510088	100.0
☑	B09510090	100.0
☑	B09510094	80.0
☑	B09510098	100.0
☑	B09510102	50.0
☑	B09510104	95.0
☑	B09510106	100.0
☑	B09510108	85.0
☑	B09510202	80.0
☑	B09510204	95.0
☑	B09510206	90.0
☑	B09510208	100.0

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	correct	ability estimation
B09310005	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	19	7.228475628
B09510002	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	10.26309262
B09510004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510006	0	1	1	1	0	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	15	4.128891681
B09510010	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17	6.349052791
B09510012	0	0	0	1	1	0	1	1	1	1	1	1	0	0	0	1	1	0	1	1	12	3.374039583
B09510016	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510018	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	18	8.418639491
B09510022	1	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	0	1	1	14	3.92651486
B09510026	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	18	7.771627679
B09510028	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	16	4.342488819
B09510030	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	11.76500885
B09510038	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	10.26309262
B09510040	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	18	7.771627679
B09510044	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	14	4.810209551
B09510046	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	18	6.349052791
B09510050	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	18	7.228475628
B09510052	0	0	0	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0	1	0	9	0.394703485
B09510054	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510056	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17	5.981935115
B09510060	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	17	5.06871674
B09510062	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	6	0.517461764
B09510064	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	17	3.92651486
B09510066	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	6.760413351
B09510074	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510076	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	17	8.418639491
B09510078	1	1	1	0	0	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	14	3.733991289
B09510084	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	12	3.374039583
B09510088	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510090	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510094	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	16	4.128891681
B09510098	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510102	1	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	1	0	10	1.890880609
B09510104	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	19	9.218247684
B09510106	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
B09510108	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	17	5.06871674
B09510202	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	16	5.347459033
B09510204	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	6.760413351
B09510206	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	18	7.228475628
B09510208	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	

Figure 3 The Proposed scorebook system

V. EXPERIMENT DISCUSSION

In the experiment, we gave no feedback to the students in class A. They could know their test grade but couldn't know why they fail with the problems. So they should find the correct answer from their textbook by their own. On the contrary, the students in class B could get all information after the pretest. Teachers make them know why they make mistake and let them find the right answers in textbook. The answer correct rate in class A and class B of pretest and posttest are shown in Figure 4 and Figure 5. Comparing the result from Figure 4 and Figure 5 that could see the progress had obvious difference. With our system's support, class B had better academic performance than class A in all of the chapters.

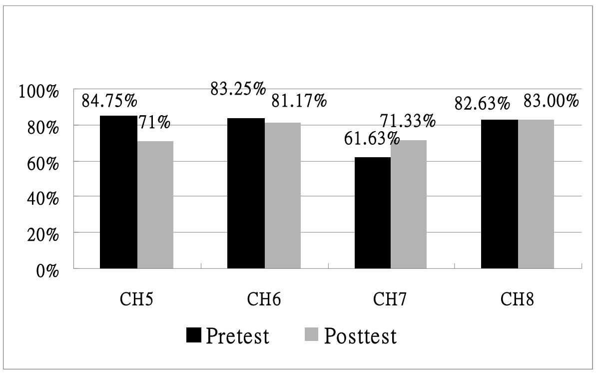


Figure 4 The answer correct rate Class A

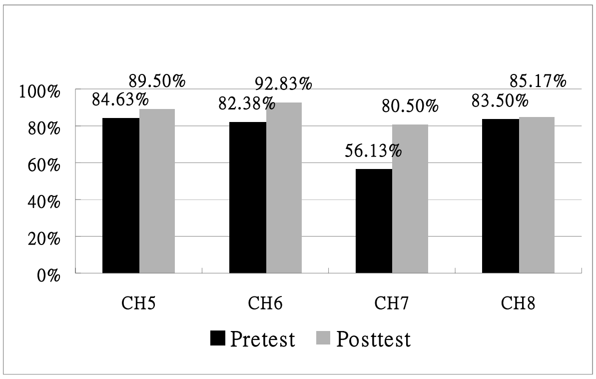


Figure 5 The answer correct rate Class B

Table 7 the pretest and posttest in class A and class B

	Class A Standard deviation		*Class B standard deviation*	
Chapter	Pretest	Posttest	Pretest	Posttest
CH5	*2.24*	*4.15*	*3.29*	*2.41*
CH6	*2.63*	*2.23*	*3.22*	*2.04*
CH7	*3.68*	*2.80*	*4.58*	*2.81*
CH8	*3.11*	*3.24*	*3.68*	*2.12*

The standard deviation is the most common measure of statistical dispersion, measuring how widely spread the values in a data set. The standard deviation is small which means many data points are close to the mean. In Table 7, the posttest standard deviation of class B is smaller than class A. In order to observe the standard deviation of two classes, Table 8 is arranged. Students who answer questions correct in class A and class B which includes the quantity of students who answer correct questions. It was apparent the class B curve looks like normal distribution than class A in Figure 6. Additionally, class B academic performance is better than class A.

Table 8 Students who answer questions correct in class A and class B

Test items	***Students who answer correct questions in class A***	***Students who answer correct questions in class B***
0~30	1	0
31~35	2	2
36~40	9	3
41~45	7	6
46~50	5	8
51~54	17	12
55~59	14	9
60	0	0

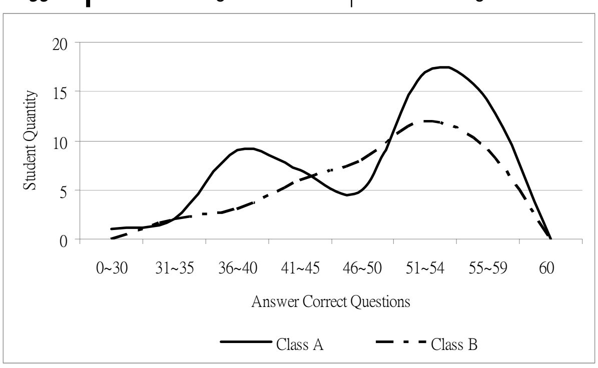


Figure 6 Students who answer questions correct disparity curves in class A and class B

According to estimated learning ability, we applied K-means to cluster 46 students into 10 groups. Table 9, Table 10 and Table 11 shows the clustered result. Table 9 shows the centers of every group are too close. This means the result is not good enough. On the opposite, Table 10 and Table 11 show s very good clustering result. Concerning the calculating complexity, this research result shows that two-parameter logistic model (IRT) is the best choice for clustering.

Table 9：One-parameter logistic model clustering

Group	Center	Item
1	4.16	1
2	4.87	4
3	3.74	2
4	2.80	6
5	3.29	7
6	1.18	6
7	2.08	7
8	1.64	7
9	2.50	4
10	0.85	2

Table 10：Two-parameter logistic model clustering

Group	Center	Item
1	1.03	7
2	26.13	2
3	15.86	1
4	5.11	7
5	6.09	2
6	7.73	5
7	9.91	3
8	2.44	10
9	3.79	8
10	12.47	1

Table 11：Three-parameter logistic model clustering

Group	Center	Item
1	0.50	7
2	25.85	2
3	10.06	2
4	4.80	7
5	5.79	2
6	7.65	6
7	15.57	1
8	2.05	10
9	3.45	8
10	12.18	1

VI. Conclusion and Future Work

Intelligent assessment technologies supports web based learning environment to provide students adaptive learning suggestions, give teachers hints to modify learning content and estimate the individual learner's ability to assist them maximize learning performance. In this paper, the methodology of estimating learning ability is discussed. We use K-means clustering method to cluster learners for small groups. We analyze the examination result of the pretest, and the posttest, and discover the difference between experimental group and control group. We apply the Item Response theory to justify each learning abilities of every examinee. A prototype web based assist system to help us to compute collected data is proposed. For the system, we could build an on-line assessment system in the front-end, and in the back-end our propose system can play an important role for the analysis in the near future. For experiment, we will analyze the pretest and posttest of one class for the experimental group and control group. Find out the learning abilities variation to apply individual learning in e-learning environment. These experiment results will provide us a valuable example for the learning management system or intelligent tutoring system construction in the future.

ACKNOWLEDGEMENTS

We would like to thank National Science Council and Chung Hua University. This research was supported in part by a grant from NSC 96-2524-S-032-001. We thank Mao-Fan Li and Tsung-Pu Lee for their hard work on this topic.

REFERENCE

[1] P. Dodds (2001a) 'Sharable Content Object Reference Model (SCORM)'. Version 1.2. Technical report, The SCORM Content Aggregation Model, Advanced Distributed Learning Initiative, October 1. Available at: http://www.adlnet.org/index.cfm?fuseaction=SCORDown

[2] W. Hodgins (2001) 'Draft Standard for Learning Object Metadata (LOM) Specification'. Proposed Draft 6.1, Institute of Electrical and Electronics Engineers, Inc. Electronic version available at http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_final_draft.pdf

[3] W. C. Chang, H. H. Hsu, T. K. Shih, and C. C. Wang, "Enhancing SCORM metadata for assessment authoring in e-learning," *Journal of Computer Assisted Learning*, Vol. 20, pp. 305-316.

[4] Nigel H. Lin, W.-C. Chang, Timothy K. Shih, and Huan-Chao Keh, "Courseware Development Using Influence Diagram Supporting e-Learning Specification," *Journal of Information Science and Engineering*, Vol. 21 , No. 5, pp.985-1005, 2005.

[5] W. C. Chang, Hui-Huang Hsu, Timothy K. Shih, and Hsuan-Che Yang, "Integrating SPC Table Formative Assessment with SCORM," *Journal of Interconnection Networks*, Vol.8, No.4, pp.305-320, 2007.

[6] Frank Baker, *The Basics of Item Response Theor*y. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.

[7] Rong-Guey Ho, Yung-Chin Yen "Design and Evaluation of an XML-Based Platform-Independent Computerized Adaptive Testing System" *IEEE TRANSACTIONS ON EDUCATION*, VOL. 48, NO. 2, MAY 2005.

[8] H. Steinhaus. *Sur la division des corp materiels en parties*. Bull. Acad. Polon. Sci., C1. III vol IV:801– 804, 1956.

[9] T.L. Kelly, The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology* 30, 17–24, 1939

[10] Dojo Tool Kits, http://dojotoolkit.org/.

Wen-Chih Chang received the B.S.B., M.S., and Ph.D. degrees from the Department of Computer Science and Information Engineering of Tamkang University in 1999, 2001 and 2005. From 2001 to 2005, he focused on SCORM and e-learning at MINELAB. His research interests include mobile agent, distance learning, cooperative learning, petri net, web technology and e-learning specifications (ADL SCORM, IMS SS and IMS QTI). He is an assistant professor in Department of Information Management, Taiwan.

Hsuan-Che Yang received his B.S., M.S. and Ph.D. degree from Department of Computer Science and Information Engineering (CSIE) at Tamkang University, Taiwan in 1996, 2005 and 2008. He is currently part-time assistant professor in the Department of educational technology at Tamkang University. His research interests include Distance Learning, Computerize Assessment, Intelligent Tutoring System and ubiquitous computing.