# A Construction Method for Personalized e-Learning System Using Dynamic Estimations of Item Parameters and Examinees' Abilities

**Yong-Sun Oh**
Department of Information Communication Engineering
Mokwon University, Daejeon, 302-729, Korea

*ABSTRACT*

This paper presents a novel method to construct a personalized e-Learning system based on dynamic estimations of item parameters and learners abilities, where the learning content objects are of the same intrinsic quality or homogeneously distributed and the estimations are carried out using IRT(Item Response Theory). The system dynamically connects the test and the corresponding learning procedures. Test results are directly applied to estimate examinee s ability and are used to modify the item parameters and the difficulties of learning content objects during the learning procedure is being operated. We define the learning unit 'Node as an amount of learning objects operated so that new parameters can be re-estimated. There are various content objects in a Node and the parameters estimated at the end of current Node are directly applied to the next Node. We offer the most appropriate learning Node for a person s ability throughout the estimation processes of IRT. As a result, this scheme improves learning efficiency in web-base e-Learning environments offering the most appropriate learning objects and items to the individual students according to their estimated abilities. This scheme can be applied to any e-Learning subject having homogeneous learning objects and unidimensional test items. In order to construct the system, we present an operation scenario using the proposed system architecture with the essential databases and agents.

***Keywords**: Personalized e-Leaning, dynamic estimation, IRT, Node, item parameter, ability parameter*

## 1. INTRODUCTION

Personalized learning has been one of the most important goals in education fields either online or offline. Therefore many researches have proposed various concepts of 'personalization in our education s world. Especially the concepts of personalized e-Learning which dynamically offers appropriate learning objects on the ground of the learner s characteristics were recently proposed in some literatures [1], [2]. But most personalized approaches consider learner preferences, interests, and habitual attitudes when analyzing their personalized characteristics for e-Learning fitting services.

On the other hand, some studies emphasize that personalized learning should consider the learner s ability or knowledge in relation to the learning subject [3]. In addition we might improve learning performance by considering learner s ability.

In the middle of 2000s, trend of personalized e-Learning has been with the logic that 'test and 'learning must be tightly related in order to offer a personalized fitting in learning processes. Moreover, e-Learning system which dynamically connects be tween 'test and 'learning in learning procedures is being one of the hottest issues so that we can obtain the highest performance in our e-Learning environment [4]. There has been many studies on the personalization for various learning environment using IRT(Item Response Theory). But these studies cannot sufficiently relates 'test and 'learning procedures dynamically with each other, and they are realistically separated so that we can hardly connect between learner s ability and item parameters. Moreover they apply 1-parameter model and discard the discrimination and the guessing parameters from references so that they can not realize a sufficient credibility of estimations.

In this paper, we propose a new scheme of personalized e-Learning system which estimates the learner s ability parameter and item parameters at the end of every Node of content objects so that we can offer a dynamic connection between 'test and 'learning procedures. In this scheme, a learner s ability parameter varies from Node to Node by its responses to the items given in the Node, and the item parameters are also varied from Node to Node by their correct/wrong answers given in the Node, all based on IRT 3-parameter model. The total content consists of several large sections. And a large section consists of many small sections etc. Minimal section consists of many Nodes. And a Node consists of SCOs(sharable content objects) the smallest conceptual elements of content object as is the case of [5]. There are two kinds of SCOs which are 'learning object SCO (obj. SCO below) and 'item SCO . The unit of re-estimation of parameters is the Node which consists of many obj. SCOs and item SCOs.

* *Corresponding author. E-mail : ysunoh@mokwon.ac.kr*
*Manuscript received May. 16, 2008 ; accepted Jun. 21, 2008*

The rest of this paper is organized as follows; Section 2 briefly describes IRT and the item response model used in this paper. Section 3 describes the system architecture and its operations which is the key section about our personalized e-Learning system. Finally Section 4 gives our concluding remarks and describes the future studies

## 2. ITEM RESPONSE MODEL

### 2.1 Logistic Model

In order to assign the most appropriate learning content objects and items for an examinee, we should have a method to measure his/her ability or knowledge about the corresponding subject and item parameters of the given items. We utilize the IRT(Item Response Theory) principle to estimate the 'ability parameter of the learner and the 'item parameters of the given items. In addition we select the 3-parameter Logistic model because our items may be multiple-choice type with 5 alternatives and have not only 'discrimination , 'difficulty but 'guessing parameters [6], [7].

Therefore the probability of correct answer by an examinee with the ability parameter $\theta$ can be written as follows:

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

where the item parameters $a$, $b$, and $c$ are discrimination, difficulty, and guessing parameters of the corresponding item respectively.

This model is a useful approximation of the Normal Ogive model. The resultant ICCs(item characteristic curves) and parameters are nearly equal in two models with an ordinary condition but the Logistic model gives us much simpler calculations in the process of parameter estimation than the Normal Ogive model. Moreover we frequently calculate the ability parameter to show an adequate courseware and estimate item parameters to renew the content information in our e-Learning operations. Henceforth our analyses should be based on the Logistic model so that we can operate our system as a real-time personalization.

### 2.2 Parameter Estimations

We distinguish two kinds of parameter estimations in this paper. One is off-line, the other is on-line estimation. We apply off-line estimation at the first time of the system setting and periodical renewals of the system. On the other hand, on-line estimation is carried out in the procedure of on-line operations of our personalized e-Learning system. We call these on-line processes as dynamic estimation of parameters. In general, off-line estimation does not need a real-time process, hence we proceed more accurate estimations using packages such as BILOG-MG with JMLE(joint maximum likelihood estimation), MMLE(marginal MLE) method etc [6]. But the on-line estimation needs a simple real-time algorithm running in a sufficiently short time ensuring our personalized e-Learning to be offered real-time in any circumstance.

The estimations mentioned above produce various kinds of results, such as item parameters, ability parameters, ICC, TCC, and True Scores for the student final assessments. In addition, we can obtain several plots about items and a histogram about learners abilities using BILOG-MG Graphics.

In order to estimate the item parameters, we give chase new discrimination($a$), difficulty($b$), and guessing($c$) parameters maximizing the following likelihood function in our on-line operations:

$$L(U \mid \theta) = \prod_{i=1}^{N} P_i(\theta)^{U_i} Q_i(\theta)^{1-U_i} \qquad (2)$$

where $U_i$ is the $i$-th item response variable having value '0 for a correct answer or '1 for a wrong answer, $N$ is the number of items manipulated in the corresponding Node. The probability $P_i(\theta)$ refers to Eq.(1) for the $i$-th item, and $Q_i(\theta) = 1 - P_i(\theta)$. In Eq.(2), all the probabilities are directly multiplied because of the local independence of items [7].

Once the item parameters are re-estimated, we consider that examinee s ability parameter can be re-estimated using these new item parameters and existing ability parameter as follows:

$$\theta_{new} = \theta_{old} + \frac{\sum_{i=1}^{N} a_i \{U_i - P_i(\theta_{old})\}}{\sum_{i=1}^{N} a_i^{\,2} P_i(\theta_{old}) Q_i(\theta_{old})} \qquad (3)$$

where $\theta_{old}$ is the ability parameter in the previous Node, and $\theta_{new}$ is the new ability parameter applied to the present Node. Originally Eq.(3) has to be used to re-estimate ability parameter when many items were tested so that the number of items $N$ may be sufficiently large. But we repeatedly apply Eq.(3) at the end of every Node to get the nearest ability parameter for an examinee. The more learning Nodes are applied, the accuracy of the ability parameter will be higher, and we can also confirm self advancement of examinee s ability in the process of re-estimations.

We offer the most appropriate learning Node for a person s ability throughout the estimation processes presented above. Selections of the best Node rely on comparison the ability parameter of the learner with the difficulty parameters of items available at the end of every Node.

## 3. SYSTEM ARCHITECTURE AND OPERATION

In this section, we present the architecture of our personalized e-Learning system and its operations. We design the framework of the system as shown in Fig.1 modifying the structure proposed in [3] and fitting to our concept of dynamic estimations of parameters in the process of personalized e-Learning. Especially, we attach off-line procedures described as 'School and 'Off-line Estimator to the conventional on-line procedures. School means the supplier of bulky item-response data. We can get an off-line estimation result by inputting the data into Off-line Estimation such as BILOG-MG.

### 3.1 Design of Content

First of all, learning content should be broken into segments

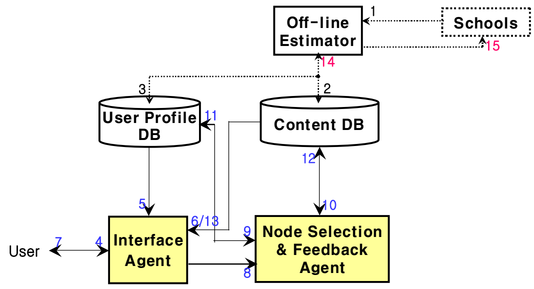


Fig.1. Framework of the personalized e-Learning System

to be used independently so that we can achieve the personalized e-Learning scheme proposed in this paper. The content segments are equivalent to the conceptual-unit object described in [5], and they have the same concept as SCO(sharable content object) in the SCORM standard [8].

The hierarchical structure of the content consists of ‘chapter , ‘large section , ‘medium section , and ‘small section etc starting from the name of subject as usual. But we invite the concept of learning ‘Node to represent the unit at which we can re-estimate the item parameters and ability parameter. Node includes many learning content SCOs and item SCOs distributed horizontally in it. Item SCO means a problem for which learner can response in a Node and content SCO is a conceptual-unit of learning object offered in the same Node. In addition, the word ‘horizontal means that there is no sequential order to study Nodes in the next higher level conceptual unit. Therefore a Node can be a ‘small section or a lower level conceptual unit.

All item SCOs in a Node have their inherent item parameters, and they are re-estimated at the end of the Node. A learning content SCOs has its own ‘difficulty parameter that is the same for directly related item SCO or average item difficulty of the Node. In this method, we can construct an efficient personalized e-Learning scheme with dynamic estimation processes by making content SCOs to possess their own difficulty parameters as items.

### 3.2 Agents and DBs

We construct agents and DBs to realize the personalized e-Learning concepts presented in this paper as shown in Fig.1. We introduce two kinds of agents and another two kinds of DBs and define their interactions.

Interface Agent(‘IF agent ) realizes HCI(human computer interface) that transmits learner ID, password, status information, and various queries. Moreover it displays content objects and items to learner and relays examinee s response to the system. On the other hand, Node Selection & Feedback Agent(‘NSF Agent ) performs parameter estimations and updates, selections of learning Nodes and offering. The key points in the process of personalized e-Learning are dynamic estimations of parameters and dynamic selections of the optimal Node. At the end of every Node, NSF agent re-estimates the examinee s ability parameter so that we can select the optimal Node available that has the nearest ‘difficulty parameter measured from the ability parameter. Although there are various methods to measure the distances from learner s ‘ability to item s ‘difficulty , we use the mean-square error criterion.

User Profile DB in Fig.1 contains learner status, accounting information, item response vector, user ability parameter, and performing information about learning Nodes. Throughout the interactions of IF agent, it transmit the resent learner s ability parameter, and then NSF agent selects the optimal learning Node available and offers to the user display by the interaction between IF agent.

Finally Content DB maintains learning content in the form of conceptual-unit objects. From the subject to the Nodes, it manages the hierarchical structure of the content objects attached by SCO numbers and their parameters. Especially learning object SCOs have ‘difficulty parameters, item SCOs have their ‘discrimination , ‘difficulty , and ‘guessing parameters. Because the amount of SCOs increases according to participations of teachers and students, management of the parameters may be more important. NSF and IF agents operate together with two DBs to accomplish personalized e-Learning procedure described in the following subsection.

### 3.3 System Operations

We present on-line and off-line procedures for the system operations to achieve the personalized learning proposed in this paper using the structure shown in Fig.1.

Off-line procedures include content authoring, initial parameter pursuits, DB initializations, and detail parameter estimation carried out periodically. It is performed with the results from off-line school separated from on-line personalized fitting operations. These off-line procedures are illustrated by broken lines in Fig.1.

Step-0 Makes learning content in the form of conceptual-unit objects and extracts the initial abilities of participants and item response vectors by off-line procedures. The item responses are extracted as direct answers correct/wrong to the item SCO presented above. The initial ability parameters are extracted from the records in the off-line tests performed up to now.

Step-1 Estimates the initial item parameters $a_i$, $b_i$, $c_i$, $i = 1, 2, 3, \ldots, N$ for all item SCOs made in Step-0 using an off-line estimator such as BILOG-MG, where $N$ is the number of item SCOs in a Node. The input data needed for this estimation are the initial ability parameters $\theta_j$, $j = 1, 2, \ldots, M$ and their item response $(100111001101000\ldots)_j$, $j = 1, 2, \ldots, M$, where $M$ is the number of examinees for the initial test. On the other hand, the difficulty parameter of learning object SCO is given by the parameter $b_i$ of the corresponding item SCO when the both are directly connected. But we give the average difficulty of the item SCOs in a Node as the difficulty parameter of learning object SCO when they are not directly connected. The connection between learning object SCO and item SCO can be decided by the authors of the SCOs.

Step-2 Saves the item SCOs with their initial parameters and learning object SCOs with their difficulty parameters in the Content DB. The specifications of Nodes, Sections, Chapters, and Subject are independently managed.

Step-3 Saves the initial learners profiles in the User Profile DB for the participants in the test operations of the system.

Now we are ready for on-line operations. The initial states of User Profile DB and Content DB given above should be used at the start point of the personalized on-line operation. However the DBs will be abundant by entering new participants and authoring new SCOs. Moreover the parameters of item SCOs and examinees abilities will be dynamically changed throughout the on-line operations. These on-line procedures are illustrated by solid lines in Fig.1.

Step-4 User login and course selection procedure. User enters the system using his ID and password via IF agent. New user has to register his personal status information before login. After login, user searches interested course, mode, and level of e-Learning scheme.

Step-5 Identifies user and transmits his ability parameter from the User Profile DB. The ability parameter of new user without any profile information will be given by '0 . This step initializes NSF and IF agent to operate for the personalized e-Learning procedure.

Step-6 Selects the optimal learning Node from the Content DB. The optimal Node is decided as a Node available with difficulty parameter nearest to the ability parameter of the user. NSF agent selects the Node that has the minimum mean-square error given below by comparing the recent ability parameter $\theta_n$ with the difficulty parameters $b_i$, $i = 1, 2, 3, \ldots, N$, where $N$ is the number of item SCOs and learning object SCOs in the Node.

$$\overline{E^2} = \sum_{i=1}^{N} \frac{1}{N} \left(\theta_n - b_i\right)^2, \quad i = 1, 2, \cdots, N \qquad (4)$$

$\theta_n$ means the initial ability parameter at this step, but it will be the recent ability after some rounds of the loop with re-estimations at the end of every Node. Therefore the examinee s ability parameter dynamically varies with the item parameters and the responses about the items in this scheme.

Step-7 Plays selected Node and obtains user s responses about the item SCOs in the Node. Displays the Node selected by NSF agent to the user and waits for his responses via IF agent.

Step-8 Transmits user s item responses to the NSF agent in order to re-estimate the ability parameter and the item parameters.

Step-9 Re-estimates the learner s ability parameter. NSF agent pursues new ability parameter by means of user s responses obtained in Step-8 and the present ability parameter from the User Profile DB using Eq.3 with Eq.1.

Step-10 Re-estimates the item parameters. NSF agent pursues new item parameters by means of user s responses obtained in Step-8 and the present parameters of SCOs from the Content DB by maximizing the Likelihood function given in Eq.2.

Step-11 Saves new user s ability parameter re-estimated above to the User Profile DB.

Step-12 Saves new item parameters for item SCOs and new difficulty parameters for learning object SCOs re-estimated above to the Content DB.

Step-13 Selects the optimal learning Node from the Content DB according to the new ability parameter changed by Step-9 and Step-11. This step is basically equivalent to Step-6 except that we utilize new ability parameter re-estimated in Step-9.

During the personalized e-Learning procedures are continued after user s login, the loop from Step-7 to Step-13 is repeated and the parameters are periodically re-estimated. The change of ability parameter of one examinee can be considered as realistic improvement of his ability about the subject, but the changes of item parameters might be considered as modifications of errors resulted from imperfect estimations executed previously. Therefore we expect that the item parameters may have little changes after several iterations of the loop.

We can fulfill off-line estimations of item parameters at the end of semester or school year for more accurate operations in the future. These processes are also depicted by broken lines in Fig.1.

Step-14 Re-estimates the item parameters and obtains various graphical results of such an operation of the system using off-line estimator BILOG-MG. We can also carry out off-line estimations for learner s ability parameters. These results are used for the initial conditions at the starting point of the next semester or school year through Step-2 and Step-3. In addition we can get ICCs(item characteristic curves) and Information Curves of items and tests, Histogram of users abilities using IRT Graphics program included in BILOG-MG. These plots are useful feedbacks for the school learning managements.

Step-15 Offers feedback information to the school. We can deliver graphical information such as ICC, TCC, IIC, TIC and measured values such as Goodness-of-fit of item parameters, Standard Error of ability estimation, True Score etc. We can use True Score as data utilized instead of conventional relative evaluation.

## 4. CONCLUDING REMARKS

We have presented a new scheme for personalized e-Learning based on dynamic estimations of item parameters and learners abilities using IRT. This scheme dynamically connects the test and the corresponding learning procedures. We have introduced a concept of learning unit 'Node including learning object SCOs and item SCOs so that new parameters can be re-estimated at the end of each Node. As a result, this scheme has improved learning efficiency in web-base e-Learning environments by offering the most appropriate learning objects and items to the individuals according to their estimated abilities. This scheme can be applied to any e-Learning subject having homogeneous learning objects and unidimensional test items.

Currently we are working toward a realization of e-Learning system for the written exam of driver s license test applying the personalization concepts proposed in this paper [9]. Especially the off-line estimations of item parameters are quite perfectly derived from the abundant data accumulated up to the present in Korea driver s license test. We expect an effective personalized e-Learning system with our high-speed web environment.

Out system needs a diversity of available Nodes at the end of each Node for more effective learning. But we suffer from the lack of content abundance and participants in the early days of system opening. We are gradually having a mature DBs in accordance with more accesses of teachers and students through Internet as time passes after system opening.

In the future, we will apply this concept to the e-Learning system with hierarchical structure of content objects to develop a personalized e-Learning scheme. We should deeply study about the 'test equating process so that we can freely deal with the test items including both multiple-choice type and subjective type on the basement of IRT.

## REFERENCES

[1] M. Balabanovic and Y. Shoham, *"Fab: Content-based, Collaborative Recommendation,"* Commun. Of the ACM, Vol. 40, No. 3, 1997, pp. 66-72.

[2] E. Leung, Q. Li, and Y. T. Zhuang, *"Media-Based Presentation with Personalization in a Web-Based eLearning System,"* Lecture Notes in Computer Science, Vol. 3818, Springer 2005, pp. 160-171.

[3] C. M. Chen, H. M. Lee, and Y. H. Chen, *"Personalized e-Learning System Using Item Response Theory,"* Computers & Education, Vol. 44, 2005, pp. 237-255.

[4] C. M. Chen and L. J. Duh, *"Personalized Web-Based Tutoring System Based on Fuzzy Item Response Theory,"* Expert Systems with Applications, Vol. 34, 2008, pp. 2298-2315.

[5] Yong-Sun Oh, *"Educational Digital Content Which Applies Conceptual Object Branch Method, and its Manipulation,"* Korea Patent No.10-0442417, 2004.

[6] Mathilda du Toit, *IRT from SSI: BILOG-MG*, Scientific Software International, Inc., 2003.

[7] F. B. Baker and S. H. Kim, *Item Response Theory – Parameter Estimation Techniques*, 2nd ed., Marcel Dekker, Inc., 2004.

[8] ADL(Advanced Distributed Learning), *SCORM 2004 3rd ed. Content Aggregation Model(CAM) Ver.1.0*, Section2.1 SCORM Content Model Components, Nov. 16, 2006.

[9] Yong-Sun Oh, *"Personalized e-Learning System for the Written Examination of Driver's License Test,"* applied for Korea Patent, June 20, 2008.

**Yong-Sun Oh**

He received B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1983, 1985, and 1992, respectively.

He worked as an R&D engineer at the System Development Division of Samsung Electronic Co. Ltd., Kiheung, Kyungki-Do, Korea, from 1984 to 1986. He joined the faculty of the Dept. of Information Communication Engineering, Mokwon University, Daejeon, Korea, in 1988. During 1998-1999 he served as a visiting professor of Korea Maritime University, Busan, Korea, where he was nominated the head of the Academic Committee of the Korea Institute of Maritime Information and Communication Sciences. Afterward, he returned to Mokwon University in 1999 and served as Dean of the Central Library and Information Center from 2000 to 2002. and as Director of the Corporation of Industrial Educational Programs from 2003 to 2005. He is currently a professor of Mokwon University and the President of Korea Contents Association. He is also working at ETRI as a Visiting Researcher.

His research interests include digital communication systems, information theory and their applications. Recently he is interested in multimedia content and personalized e-Learning.