Aberrant Learning Achievement Detection Based on Person-fit Statistics in Personalized e-Learning Systems

Ming-Tsung Liu and Pao-Ta Yu
Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan // mtlui@cs.ccu.edu.tw // csipty@cs.ccu.edu.tw

ABSTRACT
A personalized e-learning service provides learning content to fit learners’ individual differences. Learning achievements are influenced by cognitive as well as non-cognitive factors such as mood, motivation, interest, and personal styles. This paper proposes the Learning Caution Indexes (LCI) to detect aberrant learning patterns. The philosophy behind the LCI is that if any non-cognitive factor influences a learner, the effect will eventually be reflected in his/her learning achievement. Therefore, it’s our explicit attempt to build a prototype system aimed at assessing aspects of learning other than cognitive factors. This study proposes a personalized e-learning system based on Item Response Theory which considers both course difficulty and learner’s ability to provide adaptive learning paths. The LCI, which originates from the person-fit statistics in psychometric theory, statistically judges whether the observed learning achievement is significantly different from the achievement predicted by the Item Response Theory (IRT) models. If such an aberrant learning pattern is detected, a computer tutoring agent appears to notify and encourage that learner. Furthermore, human tutors may get involved periodically to offer further guidance to support learners with aberrant patterns. Experimental results show that such diagnostics could enhance the learning efficiency and smooth the learning experience.

Keywords
Intelligent tutoring systems, Personalized learning, Person-fit statistics, Item Response Theory, Learning caution indexes

Introduction
Individual differences are recognized in almost every area of performance (Waite, Wheeler, & Bromfield, 2007; Sohn, Doane, & Garrison, 2006). Educational psychologists have always emphasized the cognitive as well as the individual affective and conative differences. Human diversity is ubiquitous in education, as stated by previous researchers. Teachers and educational designers need to understand the variations in students’ attitude, motivation, and style as well as ability. Some previous studies provide personalized learning service by considering only the cognitive aspects of students. Recent studies have pointed out the important role of motivation and affectivity in cognitive activities, such as learning (Damasio, 1994; Izard, 1984).

In this study, Learning Caution Indexes (LCI) and the Item Response Theory (IRT) model together are regarded as a complete framework for personalized e-learning system. IRT models are used to estimate learner’s ability level from the learning response data, and the LCI is used to detect aberrant learning patterns by examining the fit of learning response vector to the IRT models. Note that we use dichotomous learning response in this study. A learner’s positive learning response to a course unit means that the learner can completely understand that course unit; a learner’s negative learning response means that the learner cannot completely understand that course unit. Instead of identifying all possible sources that influence learning achievements, the proposed LCI, originates from person-fit statistics, works in a different way. The philosophy behind LCI is that if any non-cognitive factors influence a learner, that effect will eventually be reflected in his/her learning achievement. We focus on the analysis of learning response. The duration, type, and frequency of various non-cognitive factors that affect learning are not the main concerns in this study.

In this paper IRT is chosen as the modeling technique because it is theoretically sound and has been widely used in many famous exams such as Graduate Record Examination (GRE) and Test Of English as a Foreign Language (TOEFL). This study applies IRT in the personalized e-learning domain, which has the following main advantages:

- **Tracking learner’s ability level**: Learner’s ability level is tracked and re-estimated after each completion of course unit.
- **Recommending course units according to learner’s ability level**: The information function inherently provided by IRT is used to recommend course units with appropriate difficulty that match learner’s ability level.
Detecting aberrant learning achievements: The Learning Caution Indexes, based on person-fit statistics, can effectively detect aberrant learning patterns in a statistically way. Detecting aberrant learning patterns is helpful to 1) encourage learners to concentrate on the learning activity, and 2) remind human tutors to pay attention to learners that may have learning difficulties.

In summary, some researchers focus on cognitive aspects when doing research in personalized e-learning domain. However, learning achievements actually result from an interaction between cognitive, affective, and conative aspects. Therefore, we attempt to build a prototype system aimed at assessing aspects of learning other than cognitive. Our research question in this study is “How to build a model to detect aberrant learning patterns?” By using the Learning Caution Indexes, aberrant learning patterns, e.g. a learner who had performed well in difficult course units but failed in some easy course units, can be detected effectively. Note that the difficulty parameters of all course units must be determined first, as described in the third section titled “Learning Diagnosis Model”. If an aberrant learning pattern is detected, the computer tutoring agent raises a warning to notify that learner and encourage him/her to concentrate. Furthermore, human tutors are involved to offer further learning guidance to support learners with aberrant learning patterns.

The remainder of this paper is organized as follows. The second section reviews the theoretical foundations related to personalized e-learning system and person-fit statistics. The third section describes the proposed learning diagnosis model. The fourth section describes the learning environment. The fifth section provides some experimental results and the sixth section draws our conclusions.

Theoretical Foundations

A Literature Review on Personalized e-Learning and Intelligent Tutoring Systems

The research background and related works are reviewed in this section. Personalized e-learning systems and intelligent tutoring systems considering cognitive, affective or conative aspects are reviewed.

Personalized e-learning systems tailor learning content for learners to fit individual differences (Graesser, Chipman, Haynes, & Olney, 2005; Hatzilygeroudis, & Prentzas, 2004; Kavcic, 2004; Litman, & Forbes-Riley, 2006). Adapting content to individual students significantly increases the speed of learning (Davidovic, Warren, & Trichina, 2003). Modeling the student’s knowledge structure is the key to support adaptive learning. Various mechanisms have been developed and evaluated by educational researchers. For example, Atolagbe, Hlupic, and Taylor (2001) discussed the role of pedagogical strategies to facilitate the acquisition of simulation modeling knowledge. Hwang (2003) proposed an adaptive learning system which provides learning suggestions by analyzing the subject materials and test results based on a concept map. Wang (2004) proposed an adaptive item selection to make students progress up the cognitive ladder based on non-symbolic neural network technology. His proposed model memorized learning paths of well-performing students, and accordingly provided personalized learning sequences for students with a similar trait level.

Chen, Lee, and Chen (2005) proposed a personalized e-learning system based on the Item Response Theory. The adaptive testing theory in computer adaptive testing (CAT) inspired them to transfer IRT into the personalized e-learning domain. This approach applies the one-parameter logistic model proposed by Georg Rasch in 1966 (Hambleton, 1985; Horward, 1990; Rasch 1966) to model the difficulty levels of course units. Furthermore, learner’s ability level could be dynamically estimated based on the maximum likelihood estimation (MLE). Their system could recommend appropriate course units to fit each learner’s ability level, but could not deal with aberrant learning patterns. The BUGGY system, which is partially similar to our proposed system, detects aberrant learning patterns in a different way (Brown & Burton, 1978). The BUGGY system uses its student model to simulate a student with “buggy” thinking. The knowledge representation model for their study is the procedural network, which breaks down a larger task into a number of related subtasks. The developers of BUGGY system attempted to enumerate the different possible procedural bugs that students might acquire while trying to solve math problems. Using a catalog of possible bugs, the BUGGY system could generate general diagnostic tests to identify learner’s mistakes. Compared with BUGGY system, our proposed system could detect aberrant learning behaviors through the IRT model without knowing anything about the learning content. Consequently, our system has fewer limitations and therefore could be applied in various domains including mathematics. Anderson and colleagues built intelligent
computer-based tutors around the Adaptive Control of Thought (ACT) theory (Anderson, Conrad, & Corbett, 1989). The basic idea was to build into the computer a model of how ACT would solve a cognitive task such as generating mathematical proofs. The tutor used the ACT theory to get the student to emulate the model. The Math tutor derived from the ACT theory and its successive versions has had a significant impact on mathematics achievements in many schools. Anderson and colleagues’ work is often cited as the most successful intelligent tutoring effort. The concept of building a cognitive model is similar to our work. However, if IRT model is used, person-fit statistics could be further integrated to detect aberrant learning behaviors.

Some educational researchers have devoted their attention to revealing how affectivity and conation influence learning. Psychologists and educators point out how the emotions affect learning (Goleman, 1995; Vygotsky, 1994). Motivated by connections between learning and student’s emotional state (Izard, 1984; Masters, Barden, & Ford, 1979; Nasby, & Yando, 1982), affective reasoning has been incorporated into computer tutoring systems to narrow the performance gap with human tutors (Conati, Chabtal, & Maclaren, 2003; Kort, Reilly, & Picard, 2001; Bhatt, Evans, & Argamon, 2004). Techniques from artificial intelligence have been studied to increase the customizability of student’s affective states in educational systems. The affective states have been modeled to recognize users’ emotions by voice, facial expressions, or even physiological signs (Kopecek, 2000; Ekman, 1999; Picard, 1997; Picard, Vyzas, & Healey, 2002).

Observable behaviors also provide implicit clues to approach student’s emotional state. Jaques and Viccari (2005) suggested that emotions can be inferred by student’s observable behavior, i.e., student’s actions in the interface of the learning environment, such as success or failure in tasks, request or refusal of help. Jaques and Viccari adopted the belief-desire-intention (BDI) model to implement the process of affective diagnosis. They used the psychological OCC model (Ortony, Clore, & Collins, 1988) that infers learners’ emotions from their actions in the system interface.

Motivation deals with the student’s desire to participate in the learning process (Ames, 1990). Motivation has been viewed as the primary determinant of student’s learning and success in school (Pintrich and Schunk, 2002). In earlier studies, motivation was seen as a personal trait, a part that depended on the genetic nature and the childhood experiences of the student (Meece & McColskey, 2001). However, more and more researchers believe that motivation is sensitive to context and can be fostered in the classroom. Based on this belief, a lot of studies have been carried out on fostering student’s motivation to learn in an educational system (Bercht & Viccari, 2000; de Vicente & Pain, 2002; del Soldato & du Boulay, 1995).

**Person-fit statistics**

The statistical technique is borrowed from the person-fit statistics in modern test theories to diagnose learner’s behavior. Person-fit statistics are developed by educational researchers to identify examinees with an aberrant test response pattern. Similarly, aberrant learning patterns of learners with learning difficulties could be identified by person-fit statistics.

Patterns with a significant number of wrong answers to “easy” questions but right answers to “hard” questions are regarded as “aberrant”. Actually, person-fit statistics are statistical methods for identifying students with aberrant patterns. Researchers have been interested in analyzing response patterns to model learners instead of using only the total score of a test (Tatsuoka, 1984). Harnisch, & Linn (1981) observed that the same number of right answers on a test could mean very different things without reference to individual characteristics. For instance, on a 20-item test, a score with 10 right answers can be obtained in 184,756 different ways.

The detection of aberrant examinees using person-fit statistics inspired us to transfer person-fit statistics into the personalized e-learning domain for learning diagnosis. Learners are diagnosed according to their observable behaviors in this study, as suggested by Jaques and Viccari (2005). There are several robust person-fit statistics, such as the Person-fit U3 statistic (Van der Flier, 1977, 1982) and the Caution Index (Sato, 1975, 1982). Wright and his students proposed two person-fit statistics (Wright, 1977; Wright & Stone, 1979) INFIT and OUTFIT, whose distributional properties have been exhaustively investigated and reported by Smith (1986). Most person-fit statistics can be employed to diagnose the learning response data after necessary modifications. However, INFIT has been shown to be near optimal in identifying spurious scores at the ability distribution tails (Rudner, 1977). Thus INFIT...
and OUTFIT are chosen as a sample implementation to demonstrate that person-fit statistics could be applied to recognize aberrant learning patterns.

**Learning Diagnosis Model**

**Determine the difficulty parameters of course units**

In the learning diagnosis model, the one-parameter item characteristic function is used. The difficulty parameters of course units must be determined first. Although the difficulty parameters can be assigned manually by course experts, we employ the following procedure to estimate them systematically. Course experts identified learning concepts and then design a corresponding test item for each learning concept. These test items are then taken by a number of examinees, and the testing results are analyzed by BILOG-MG (Mislevy & Bock, 1990) to obtain the appropriate difficulty parameter of every test item. BILOG-MG is a software program for IRT analysis of dichotomous (correct/incorrect) data, including fit and differential item functioning. Course units are then designed to cover all learning concepts. Since the test items and the course units are both derived from the learning concepts, each course unit has the same difficulty parameter value as its corresponding test item. This approach can obtain more dependable difficulty parameters than manual calibration, because parameters are estimated based on real experimental data (Chen, Lee, & Chen, 2005; Chen, Liu, & Chang, 2006).

**Estimate learner’s ability level**

Maximum likelihood estimation (MLE) (Horward, 1990) is a popular statistical method to fit a model to data and to provide estimation of parameters in that model. In this study, MLE is applied to estimate learner’s ability level. We first assume that a randomly chosen learner responds to a set of \( n \) course units with response pattern \( (U_1, U_2, \ldots, U_n) \), where \( U_j (j=1, 2, \ldots, n) \) is either 1 or 0 for the \( j \)th course unit. \( U_j = 1 \) means that learners can completely understand the \( j \)th course unit, and \( U_j = 0 \) means that learners cannot completely understand the course unit. Based on the assumption of local independence (Baker & Frank, 1992; Wang, 1995), the formula which estimates learner’s ability level based on the difficulty parameters of course units is:

\[
L(U_1, U_2, \ldots, U_n|\theta) = \prod_{j=1}^{n} P_j(\theta)^{U_j} Q_j(\theta)^{1-U_j}
\]

where \( P_j(\theta) = \frac{e^{\theta b_j}}{1 + e^{\theta b_j}} \) and \( Q_j(\theta) = 1 - P_j(\theta) \). Note that \( P_j(\theta) \) denotes the probability that a learner with ability level \( \theta \) can completely understand the \( j \)th course unit, \( b_j \) is the difficulty parameter of the \( j \)th course unit. Since \( P_j(\theta) \) and \( Q_j(\theta) \) are the functions of learner’s ability level \( \theta \) and the difficulty parameters of course units, the likelihood function \( L(U_1, U_2, \ldots, U_n|\theta) \) is also a function of these parameters. Learner’s ability level \( \theta \) can be estimated by computing the maximum value of the likelihood function (Hambleton, Swaminathan, & Rogers, 1991). The method of the maximum likelihood estimation requires two input parameters to evaluate learner’s ability level: the difficulty parameters of the course units and the yes/no responses from learners to course units.

Ability levels are limited between \(-3\) and \(3\). Learners’ ability levels are estimated based on their feedback. If a learner can completely understand the content of the recommended course units, his/her ability level will be increased; otherwise, his/her ability level will be decreased. The estimated ability levels are further used by the information function in the course recommendation procedure to rank course units.

**Recommend a course unit**

In this study, course recommendation considers both course difficulty and learner’s ability, which affect both learners’ interest and learning results. Two common approaches, the maximum information strategy and the Bayesian strategy (Baker, & Frank, 1992; Hambleton, 1985; Hulin, Drasgow, & Parsons, 1983), are available to choose appropriate items for examinees in modern test theories. Since the Bayesian strategy is more complicated than the maximum information strategy, the latter is employed to recommend appropriate course units. The
maximum information strategy chooses a course unit with the difficulty parameter that exhibits the maximum information value for a learner with ability level \( \theta \). The information function is defined as:

\[
I_j(\theta) = \frac{(D)^2}{\left[e^{D(\theta-b_j)}\right]^2[1+e^{-D(\theta-b_j)}]^2}
\]  

(2)

where \( I_j(\theta) \) is the information value of the \( j \)th course unit for a learner with ability level \( \theta \), and \( b_j \) is the difficulty parameter of the \( j \)th course unit. \( D \) is a scaling constant often set to \( D = 1.702 \) when the logistic function in formula (2) needs to approximate a cumulative normal probability function. When “normal ogive” scaling is used (i.e., \( D = 1.702 \)), and assuming that \( \theta \) is normally distributed, we can treat \( \theta \) as a z-score. Normal ogive scaling therefore facilitates basic interpretations of \( \theta \) for practitioners familiar with z-scores and the normal curve.

After calculating the information value of every course unit, the course-recommendation module gives the recommendation list of course units, which are sorted by the ranking order of the information function value. Learners then choose a course unit to study from the recommendation list. Our system records each learner’s learning history and feedback responses into the user profile database. The learning history recorded in the user profile database is then used to detect aberrant learning patterns as described in the next subsection.

**Detect aberrant learning achievements**

The proposed Learning Caution Indexes: Learning_INFIT and Learning_OUTFIT are delivered from the INFIT and OUTFIT statistics respectively. IRT-based person-fit statistics are designed to evaluate the misfit of an observed response vector to IRT model by calculating the probabilities associated with student’s ability parameter and item parameters. According to the IRT model, if the probability of a correct response from a student is high, the hypothesis is posited that the student should answer that item correctly, and vice versa. INFIT is an information-weighted sum and OUTFIT is based on the conventional mean of squared standardized residuals (Bond & Fox, 2007). Learning_INFIT is defined as:

\[
\text{Learning\_INFIT} = \frac{\sum_{j=1}^{n} [U_j - P_j(\theta)]^2}{\sum_{j=1}^{n} P_j(\theta)Q_j(\theta)}
\]

(3)

and Learning_OUTFIT is defined as:

\[
\text{Learning\_OUTFIT} = \frac{\sum_{j=1}^{n} P_j(\theta)Q_j(\theta)}{\sum_{j=1}^{n} [U_j - P_j(\theta)]^2}
\]

(4)

where \([U_j - P_j(\theta)]\) is the residual (\(U_j\) is regarded as an observed response and \(P_j(\theta)\) is regarded as the expected response value), \(P_j(\theta)Q_j(\theta)\) is the statistical information (model variance) according to Fisher information theory (Pratt, 1976), and \([U_j - P_j(\theta)]^2\) is the squared standardized residual.

The behavior of Learning_INFIT and Learning_OUTFIT is demonstrated by some response patterns in Table 1. Four typical response patterns, which are modeled, carelessness, lucky-guessing, and special-knowledge, are listed (Linacre & Wright, 1994). The learning pattern represents learner’s learning outcome of sixteen course units which are sorted by their difficulty. ‘1’ response means that the learner can completely understand the course unit and ‘0’ response means that the learner cannot completely understand the course unit. The modeled pattern may be regarded as a normal pattern. However, the carelessness pattern may be interpreted as a possible aberrant learning pattern that a learner succeeded in difficult course units but failed an easy course unit. Similarly, the lucky-guessing pattern may be interpreted as another possible aberrant learning pattern that a learner surprisingly succeeded in a difficult course unit that is far beyond his/her ability level. The special-knowledge pattern may be interpreted as a possible aberrant learning pattern that a learner succeeded in some difficult course units but failed some easy course units that s/he could master. This pattern may have a slightly different interpretation in personalized e-learning domain. In a test, examinees answer test items according to their own knowledge. Wrong responses may be made to test items with unknown concepts, even though these concepts are very easy. On the contrary, learners have the chance to learn
these easy concepts from the course units, but they still failed. In our experiences, this pattern might suggest that either learners have individual interests in some specific concepts, or there are some flaws in the course unit so that learners can not understand these learning concepts. Learning_OUTFIT is sensitive to carelessness and lucky-guessing pattern, and Learning_INFIT is sensitive to special-knowledge pattern.

<table>
<thead>
<tr>
<th>Learning Pattern</th>
<th>Diagnosis</th>
<th>Learning_INFIT</th>
<th>Learning_OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy: 0110110100</td>
<td>Modeled</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Medium: 1111110000</td>
<td>Carelessness</td>
<td>1.0</td>
<td>3.8*</td>
</tr>
<tr>
<td>Hard: 1111000000</td>
<td>Lucky-guessing</td>
<td>1.0</td>
<td>3.8*</td>
</tr>
<tr>
<td>Hard: 1110001110</td>
<td>Special-knowledge</td>
<td>1.3*</td>
<td>0.9</td>
</tr>
</tbody>
</table>

(* marks high Learning_INFIT/Learning_OUTFIT values which indicate aberrant learning responses)

Each time when a learner finishes a recommended course and posts a learning response, his/her learning response pattern is diagnosed by the Learning Caution Indexes (Learning_INFIT and Learning_OUTFIT). The thresholds are both 1.20 which rejects response strings manifesting more than 20% unmodeled noise. A high value (greater than the threshold) of Learning_INFIT or Learning_OUTFIT is interpreted as that the learner is possibly learning in a perfunctory or careless manner. Once such an aberrant learning pattern is detected, a computer tutoring agent appears to notify and encourage that learner.

The Learning Environment

Sample courseware

A sample courseware “Library Tutorial” was designed in this study. Several experienced librarians were invited as course experts to analyze the learning concepts. “Library Tutorial” is aimed at helping students who are not familiar with the library system. The prototype of the personalized e-learning system with learning diagnosis was implemented to evaluate the performance of the Learning Caution Indexes. A total of 375 freshmen in National Chung-Cheng University, Chia-Yi, Taiwan took the exam with 40 test items that cover all learning concepts. The difficulty parameters of course units were estimated according to the analysis of the testing result using BILOG-MG.

![User interface of the proposed learning system with course unit recommendation](image)
Personalized e-learning system

Learner’s current ability level, the difficulty of current course unit, and the values of Learning Caution Indexes are all displayed in the user interface of the personalized learning environment as shown in Figure 1.

Each time when completing a course unit, the learner was asked to answer the following question in the learning questionnaire: “Do you completely understand this course unit?” as shown in Figure 2. In this study, we always use the ability level estimated from the result of the self-rating learning questionnaire. The sample course “Library Tutorial” was not a regular course, and the learners did not have the pressure to pass this course. Moreover, learners were reminded and recommended to give honest answers to the learning questionnaire. Therefore, we might reasonably assume that learners provide reliable answers on whether they understood the unit or not. However, alternative mechanisms such as “a test item after finishing each course unit” could be used instead to prevent the problem that if a learner thinks that s/he understands completely a course unit (“yes” answer) when, in fact, s/he does not.

![Figure 2. Learning questionnaire page which collect learning response for a course unit](image)

![Figure 3. An agent appears to notify and encourage a learner with an aberrant learning pattern](image)
After answering the question “Do you completely understand the course unit?”, learner’s ability level is re-estimated and Learning Caution Indexes are applied to detect aberrant learning patterns. A computer tutoring agent appears to notify and encourage that learner to concentrate on learning if an aberrant learning pattern is detected as shown in Figure 3.

The system architecture of the personalized e-learning system with learning diagnosis is illustrated in Figure 4.

![System Architecture Diagram]

*Figure 4. System architecture (The labeled numbers indicate the procedure of system operations.)*

Based on the system architecture, the details of the system operations are described as follows.

**Step 1** Authenticate a learner’s identification when receiving a login request through the Internet.

**Step 2** Retrieve the learner’s learning portfolio from the user profile database.

**Step 3** Load the difficulty levels of course units that have not been studied from the course database.

**Step 4** Load the learner’s ability level from the user profile database. A moderate ability level is assigned to new learners after their first login.

**Step 5** Recommend a list of course units. Course units are ranked based on the degree of information they provide to the learner’s current ability level.

**Step 6** Fetch and launch the selected course unit from the course database.

**Step 7** Deliver the selected course unit to the client via the Internet.

**Step 8** Deliver the learning questionnaire: “Do you completely understand this course unit?” and collect the feedback after the learner finishes a course unit.

**Step 9** Load the learning feedbacks of the learner from the user profile database.

**Step 10** Re-estimate the learner’s ability based on the learning feedbacks (completely understand recommended course unit or not).

**Step 11** Update the learner’s ability in the user profile database.

**Step 12** Diagnose the learning behavior by using the Learning Caution Indexes to detect aberrant learning patterns.

**Step 13** Record the result of diagnosis to the user profile database.

**Step 14** Activate the computer tutoring agent to encourage a learner with a detected aberrant learning pattern.

Repeat Steps 3 to 14 until the learner finishes all course units. This process may be suspended by learner’s logging out of the system and be resumed at learner’s next login. The flowchart of the personalized e-learning system from the user’s view is illustrated in Figure 5.
Evaluations and Results

Experimental setup procedure

The experimental procedure was carried out as follows. The main differences between the experimental group and control group were the course recommendation, learning diagnosis, and the targets of periodical learning guidance from human tutors, as shown in Table 2. Learners in experimental group took both additional course recommendation by the IRT model and learning diagnosis by the Learning Caution Indexes. In the experimental group, human tutors gave learning guidance to learners with detected aberrant learning patterns; in the control group, human tutors gave learning guidance to randomly chosen learners because the learning guidance from human tutors is a control variable and therefore must be given in both group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Course recommendation</th>
<th>Learning diagnosis</th>
<th>Human tutoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>IRT model</td>
<td>LCI</td>
<td>Periodical learning guidance to learners with detected aberrant learning patterns</td>
</tr>
<tr>
<td>Control group</td>
<td>None</td>
<td>None</td>
<td>Periodical learning guidance to randomly chosen learners</td>
</tr>
</tbody>
</table>

A total of 62 freshmen (other from these 375 freshmen who had participated in the exam for estimating difficulty parameters of course units) enrolled in the “Library Tutorial”. The students were randomly divided into experimental group and control groups. After learners had logged in for the first time, they were asked to complete a pretest with 20 test items (both in experimental group and control group); after completing all course units in the “Library Tutorial”, the learners were asked to take another 20-item post-test to evaluate their learning achievements.
Evaluation of the performance

The one-way Analysis of Covariance (ANCOVA) was performed to test the difference of achievements between the experimental group and the control group. Since the pretest had been administered, the scores in the pretest were used as the covariate in the analysis. The scores in the post-test were used as the dependent variable.

The descriptive statistics for the ANCOVA analysis are depicted in Table 3, whereas Table 4 presents a summary result of the ANCOVA analysis on the overall post-practice achievement test. The ANCOVA results ($F=5.14$, $P=0.027$) indicate that the experimental group scored significantly higher than the control group in the post-test. It can be concluded from the experimental results that the Learning Caution Indexes can significantly improve the learning achievements for students who used the proposed personalized e-learning system. In table 3, although it seems that learners from the control group had very little improvement of knowledge during the learning process (their mean of scores only increased from 47.74 to 51.77). However, the standard deviation of the post-test scores in the control group (25.35) is much higher than the standard deviations in pretest (16.58) and in experimental group (20.47). We believe that learners with aberrant learning patterns need some proper support; otherwise, they might soon give up learning and get poor learning achievements which not only decrease the mean but also increase the standard deviation of post-test scores as in the control group.

| Table 3. Descriptive statistics of the scores of the pretest and posttests |
|---------------------------------|---|---|---|---|
| Group                          | N | Post-test scores | Pretest scores (covariant) |
|                                |   | Mean  | SD   | Mean  | SD   |
| Experimental Group             | 31 | 65.97 | 20.47 | 43.23 | 19.38 |
| Control Group                  | 31 | 51.77 | 25.35 | 47.74 | 16.58 |

| Table 4. Summary results of the ANCOVA analysis on the posttest scores |
|-----------------------------|---------|-----|-----|-----|
| Source                      | SS      | d.f. | MS  | F    | P    |
| Between Group               | 2713.24 | 1   | 2713.24 | 5.14* | .027 |
| Error                       | 31148.90 | 59  | 527.95 |

*P<0.05

In our experiment, totally 33 aberrant learning patterns (in which 42% were reported by Learning_OUTFIT and 58% were reported by Learning_INFIT) were detected in the experimental group. 15 learners finished all course units without any detected aberrant patterns. 5, 7, 2, and 2 learners had 1, 2, 3, and 4 detected aberrant patterns respectively.

Evaluation of the degree of satisfaction

Questionnaires were applied to evaluate the learners’ degree of satisfaction in the experimental group. We collected learners’ responses to determine whether the recommended course materials meet most learners’ requirements. The proposed system collected learners’ responses to the questions: “Do you understand the content of the course material?” and “How do you think about the difficulty of the recommended course material?”. The result listed in Table 5 seems to show that most recommended course materials were moderately difficult. We may interpret the results as that the proposed system recommended suitable course units to learners.

| Table 5. Learners’ degree of satisfaction for the recommended course materials |
|---------------------------------|---------|-----|-----|-----|-----|-----|
| Question                        | Answer: Learners’ choices | Yes | No  | Total |
| (a) Do you understand the content of the recommended course materials? | Very easy | Easy | Moderate | Hard | Very hard | 1240 |
|                                | 83 (7%) | 353 (28%) | 664 (54%) | 103 (8%) | 37 (3%) | 1240 |
| (b) How do you think about the difficulty of the recommended course material? | 1062 (86%) | 178 (14%) | 1240 |
Furthermore, learners were asked to fill in three designed questionnaires after s/he finished the whole learning process. The five-point Likert scale was applied to evaluate the degree of satisfaction with the proposed system. Table 6 displays learners’ responses. Experimental result shows that the personalized service of the proposed system is quite acceptable for learners.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer: Learners’ choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) How do you feel that the course materials recommended by our system are appropriate?</td>
<td>Very suitable</td>
</tr>
<tr>
<td></td>
<td>5 (16.1%)</td>
</tr>
<tr>
<td>(b) Do the personalized services provided by our system satisfy your requirement?</td>
<td>Very satisfactory</td>
</tr>
<tr>
<td></td>
<td>6 (19.4%)</td>
</tr>
<tr>
<td>(c) Does the learning diagnosis provided by our system satisfy your requirement?</td>
<td>Very satisfactory</td>
</tr>
<tr>
<td></td>
<td>3 (9.7%)</td>
</tr>
</tbody>
</table>

**Empirical cases from experimental data**

Two empirical cases are discussed to demonstrate the capability of the LCI for detecting aberrant learning patterns. The course units in these two cases were presented in different orders because different course units were recommended by our personalized e-learning system according to each learner’s abilities.

Learning_INFIT and Learning_OUTFIT often exceeded their thresholds in the initial phase of ability level estimation. In other words, Learning_OUTFIT and Learning_INFIT were not stable in the first few course units, because the estimation of learner’s ability level was not reliable due to insufficient sample size of learning outcomes. However, both the Learning Caution Indexes and learner’s ability level became stable and reliable with sufficient sample size (after 3 to 6 course units). In this study, all false alarms in initial phase were ignored by our system.

In the first case shown in Figure 6, learner’s ability level became stable after initial phase but had a sudden drop after the fifteenth course unit. At the same time, the value of Learning_OUTFIT, which is sensitive to carelessness/sleeping pattern, exceeded the threshold (1.72) and our system raised a warning about that aberrant learning pattern. The value of Learning_INFIT also increased but did not exceed the threshold (1.16).

![Figure 6](image.png)

*Figure 6. Detection of aberrant learning patterns using LCI: A learner with sleeping/careless pattern. The y-axis represents the values of ability level, Learning_INFIT, and Learning_OUTFIT respectively.*
Aberrant learning patterns are not necessarily accompanied by sudden changes of learner’s ability level. However, the LCI has the capability to detect precisely such aberrant learning patterns without sudden changes of learner’s ability level. In the second case, the learner performed well (the ‘1’ responses) in course units with difficulty level 2.34, 2.27, 2.29, 2.28, and 2.31 but failed (the ‘0’ responses) in some easy course units with difficulty level 2.09, 2.13, and 2.14 as shown in Figure 7. However, learner’s ability level was surprisingly stable after the initial phase. In such cases, aberrant learning patterns could still be identified. Learning_INFIT, which is sensitive to special-knowledge pattern, detected the aberrant learning pattern after the ninth and the thirteenth course units. The thresholds of Learning_INFIT and Learning_OUTFIT are both 1.20 which accepts response strings manifesting fewer than 20% unmodeled noise. Due to the 20% noise-tolerance capability of LCI, aberrant learning patterns are detected after accumulating enough number of ‘1’ responses in some difficult course units and ‘0’ responses in some other easy course units.

![Figure 7. Detection of aberrant learning patterns using LCI: A learner performs well in difficult course units but fails in easy course units (special-knowledge pattern)](image)

**Conclusions**

The Learning Caution Indexes were proposed in this study to detect aberrant learning patterns by comparing the expected learning achievement with the observed learning achievement. We proposed a mechanism to diagnose whether the personal state of a learner is appropriate for learning by analyzing their explicit learning feedback collected from a learning questionnaire. Our system raises a warning to notify the learner who has a possible aberrant learning pattern. Furthermore, human tutors are involved to offer further learning guidance to support learners with aberrant learning patterns. Experimental results exhibit the effectiveness of the personalized e-learning system with the Learning Caution Indexes.

**References**


