# Quantifying the Learning Efficiency of Programming Learning and Its Implications

Yuh-Huei Shyu

**Abstract—*With the advent of distance learning, educators are having a hard time deciding whether to adopt this kind of education model or not. We need some metric systems to judge the superiority of different education models. To simplify the problem, our discussion will be restricted to programming language courses. Focusing on the issue of "learning efficiency," in this paper, we will define some quantifiable parameters and measuring rules. These parameter values can reflect the features of the underlying education models. Derived parameter values not only reveal bottlenecks of the current programming learning processes, but also provide guidelines for designing better education models. The salient results are: (1). the learning inefficiency is in fact an intrinsic property of the traditional model itself, and (2). "Active Teaching and Passive Learning for Novices." is a better teaching strategy for novices. Finally, combining all results we obtained, we propose the "CD-type learning model." This model might provide a very efficient learning environment for novice programmers.***

**Index Terms—*learning efficiency, Distance learning, Active learning, Passive teaching, Passive learning, Active teaching, CD-type learning***

## 1. INTRODUCTION

WITH advances of Internet technologies, hundreds of distance learning systems have been designed, implemented, and adopted. Distance learning systems provide environments that break temporal and spatial limitations [2,3,4]. Some researches indicate that distance learning is or can be as effective as traditional education [4,9]. While the latest report given in SIGCSE'04 indicates that online students are less successful than on ground students [11]. These reports bring the following question to the forefront:

> *"Is distance learning superior to the traditional education system?"*

This question is very hard to answer. The major difficulty comes from the fact that the meaning of "superiority" is not clearly defined. Current methods for judging the superiority of different education models are mainly based on learners' final scores or questionnaires. This kind of comparison seems to be very subjective. Thus, we need some metric systems to measure the performances of education models.

Typical education activities involve not only the *instructors* and *learners*, but also different *underlying education models*. We found that the underlying education model plays a key role on learners' learning efficiency. Once an instructor adopts an education model, he also sets the minimum time requirement for learners to master the designated course. We will discuss this issue later in the paper.

In this paper, the traditional education model will be defined as teaching activities that using no Internet technologies or mailing mechanisms; while the distance learning model is defined as teaching activities that using Internet technologies. In the traditional education model, instructors adopt different types of teaching methods such as lectures, discussions, and hands-

Manuscript received July 24, 2004.

Y. Shyu is with the Department of Information Engineering, Tamkang University, Taiwan, R.O.C., (e-mail: yuhhuei@mail.tku.edu.tw).

on. From this point forward, we will use the words "traditional education models" to represent different types of the traditional education model. Analogously, we will use the words "distance learning models" to represent different types of the distance learning model. First, let's point out the following facts.

1. A course teaching activity involves at least three elements: the instructor, the learners, and the underlying education model.
2. Elements of the instructor and the learners are changeable.
3. Most learners' final scores highly depend on the instructor's teaching quality.

Using learners' scores to judge the superiority of an education model is very subjective and has the following shortcomings.

1. Little or no consideration of the learner's background, talent, and motivation.
2. Little or no consideration of the instructor's teaching quality.
3. Little or no consideration of the influence of the underlying education model.

We need a clear definition of the meaning of "superiority". To achieve that, we need to decide parameters and measuring rules so that selected parameters will reflect the superiority of the designated education model as precisely as possible. In addition, these parameter values should be quantifiable. In this paper, we will focus only on programming courses without further specification. These parameter values can reflect the features of the underlying education models. Derived parameter values not only reveal bottlenecks of the current programming learning processes, but also provide guidelines for designing better education models. The salient results are: (1). the learning inefficiency is in fact an intrinsic property of the traditional model itself, and (2). "Active Teaching and Passive Learning for Novices." is a better teaching strategy for novices. Finally, combining all results we obtained, we propose the "CD-type learning model." This model might provide a very efficient learning environment for novice programmers.

The remainder of this paper is organized as follows. In section 2, we will discuss the concept of "learning efficiency" and define some measuring parameters and rules. In section 3, we will evaluate parameters for different models. In section 4, we will discuss issues related to comparison of education models. In section 5, we will propose a new education model called "CD-Type learning model." In section 6, we will provide conclusions and recommendations for future research.

## 2. LEARNING EFFICIENCY

### A. Definitions

There are many analogous definitions for the words "effective" and "efficient" in dictionaries and reports, we will use the definitions given in [1]:

1. Effective means doing things successfully.
2. Efficient means using resources wisely and without unnecessary waste.

Note that whenever we talk about effectiveness we often use some subjective criteria. While when we talk about efficiency, the parameter "time" is considered explicitly. Now, we will extend the above definitions to course teaching and learning.

1. "Teaching effectively" means teaching a course successfully.
2. "Learning effectively" means learning course content successfully.
3. "Teaching efficiently" means using teaching resources available wisely and without wasting unnecessary time.
4. "Learning efficiently" means using learning resources available wisely and without wasting unnecessary time.

To judge the superiority of an education model, it would be more convincing to judge its efficiency, rather than its effectiveness, since learners' grades only reflect the teaching/learning effectiveness under a pre-scheduled syllabus with a designated instructor. These data should not be used to determine the superiority of the underlying education model.

*B. Parameters of Learning Efficiency*

In general, a course includes many students and one instructor. Therefore, instead of considering "teaching efficiency," we shall consider the "learning efficiency" of a course. We will use the time required to reflect the status of learning efficiency. Learning efficiency is influenced by three major factors: the instructor, the learners, and the education model used. The first two factors are changeable (subjective) parameters; therefore, we shall ignore these two factors. The third factor is what we will try to quantify.

The concept of efficiency takes two values to compare: the base value and the target value. We will use the concept of span of total time scheduled as the base value, and the concept of span of total time actually used as the target value. For example, we define one semester as the base value for a given course. The total time that a learner needs to "finish" this course is defined as the target value. Here, we will let the meaning of "finish" be open. From the instructor's aspect, it may mean when a learner passes the course. While from a learner's aspect, it may mean when he is confident with the course content.

The learning efficiency depends on both the base value and the target value. These values have the following interpretations.

1. Base value depends on the education model.
2. A smaller base value means that the course is using a more efficient model.
3. Target value depends on both the learner and the instructor.
4. A smaller target value means that the teaching is more effective.

We will now provide assumptions and definitions that will be used in later discussions.

Assumption 1:

*Self-study of a course takes more time than being guided by a qualified instructor.*

Definition 1: Course Content Delivered Time ($T_{ccd}$)

*The "Course Content Delivered Time" of a given course is defined as the time interval between the beginning of the first scheduled lecture and the end of the last scheduled lecture. In the traditional education model, the $T_{ccd}$ of a given course is about a semester or a quarter.*

Definition 2: Maximum Lectures Interval Time ($T_{mli}$)

*The "Maximum Lectures Interval Time" of a given course is defined as the maximum time interval between two adjacent scheduled lectures.*

Definition 3: Course Mastering Time ($T_{cm}$)

*The "Course Mastering Time" for a learner in a given course is defined as the time interval from when he starts the course to the time when he "thinks" that he has mastered the course content.*

We shall leave the meaning of "thinks" stated above open. Intuitively, the lower value of $T_{cm}$ means that the learning is more efficient.

Assumption 2:

*Under the same definitions given above, we will leave it for related people to decide if the learning is efficient.*

Assumption 3:

*Assume that most students study the course according to the pace of the instructor's lectures or course syllabus*.

The whole idea here is trying to transfer the abstract concept of learning efficiency to some measurable quantities. These kinds of parameters may not be as precise as those used in other research. However, they are quite enough to reveal bottlenecks of learning efficiency. We have noticed following important results:

1. $T_{ccd}$ sets a maximum time limit for delivering a course.
2. $T_{ccd}$ also sets a minimum time limit for mastering course content.
3. $T_{ccd}$ and $T_{mli}$ values depend on the education model used and have nothing to do with the instructor's teaching quality.
4. $T_{cm}$ depends on both $T_{ccd}$ and the instructor's teaching quality.
5. Under the above assumptions, for a given learner and a course, $T_{cm}$ ? $T_{ccd}$

*C. Implications of $T_{ccd}$, $T_{cm}$ and $T_{mli}$ Values*

There may be many reasons for why a learner cannot learn efficiently. Most people

would agree that the quality of an instructor plays a dominant factor in learners' learning efficiency in a given course. Nevertheless, previous results indicate that $T_{ccd}$ plays a part in determining the minimum time required to master the course content. Since that $T_{ccd}$ value has nothing to do with the teaching qualities of the course instructor, the following result is surprising:

> *"Learning inefficiency is in fact an intrinsic property of the traditional education model."*

Assigning a well-qualified instructor to a course is very important. However, since the value of $T_{ccd}$ is bound, it seems that the traditional education system has no workable way to bypass this barrier. Also, notice that if we want to design a more efficient learning model, then we have to minimize both the $T_{ccd}$ and $T_{cm}$ values.

Some learners learn things fast, while others learn slowly. Remember that the third assumption states that most students study the course according to the pace of their instructor. Therefore, the $T_{mli}$ value is a factor in the speed of learning. In general, a bigger $T_{mli}$ value means that the learning pace is slower, and this will inevitably result in the slowing down of the progress of a fast learner. Therefore, we need a new learning environment that could provide a "true" self-pace learning.

To develop such a new learning environment, we need new concepts, new pedagogy theories, and new models. An easier way to think about the solution of this problem is trying to reduce the $T_{ccd}$ and $T_{mli}$ values while still providing good content.

## 3. $T_{CD}$, $T_{MLI}$ AND $T_{CM}$ VALUES OF DIFFERENT MODELS

### A. Traditional Education Model

In traditional programming language courses, values of $T_{ccd}$ and $T_{mli}$ are very easy to compute. Since the value of $T_{cm}$ depends on the individual (educator or learner), it is not as easy to get its exact value. However, the minimum value for average learners is very easy to get (e.g., using the criterion of passing the course).

### B Distance Learning Model

Distance learning models do not provide a uniform learning pattern. Therefore, its $T_{ccd}$, $T_{mli}$ and $T_{cm}$ values are highly depend on the course arrangement and the educator's teaching style.

While most distance learning courses provide text-only content, some courses do provide video content. Video content often leads to lengthy delay. Since that the distance learning model is very sensitive to multimedia data's delay, therefore, above parameters are not adequate to reflect its efficiency. More researches need to be carried out.

## 4. ISSUES RELATED TO MODELS' COMPARISONS

### A. Comparing Different Education Models

The education models just provide bare platforms to carry out the teaching activities. This leaves instructors a great flexibility to decide what kind of strategy is to be used. Associating different courses and teaching strategies with different education models results in a large number of possible combinations.

Different courses carry different characteristics. Therefore, instructors use different models and strategies to carry out the teaching activities, such as lecture, discussion, or self-study. Quantifiable parameters introduced above reflect some time limits of the underlying models. Yet, we still don't have the measuring rules to measure those changeable factors. Therefore, any kind of comparison will be criticized as incomplete and subjective.

### B Appreciation Instead of Comparison

In order to avoid the incomplete comparisons, we use the method of "appreciation," instead of the method of "direct comparison." This method goes like this: we first raise some important issues and then discuss them. Through discussion, we will have a deeper understanding about the problems involved. *At the end of this section, no conclusions will be made. Instead, we will leave readers to judge which model is better.*

### *C. For a Novice, Which One Is More Important: Direct Interaction or Demo Together With Imitation?*

> "In the beginning of the newborn baby's learning process, which one is more important, direct interaction or demo together with imitation?"

Let's start by looking at the typical growing process of a child. After a baby is born, he spends lots of his time listening and observing Mom's demonstrations (this is a one way communication). In addition, he tries very hard to imitate Mom's voice, gestures, and movements. It takes quite a long time to have a "meaningful" interaction with Mom and other people. During these years, he will build up his fundamental linguistic and physical capabilities.

"Do newborn babies ask meaningful questions?" The answer is definitely "No"; the baby just keeps on observing and imitating. These facts lead us to reconsider the necessity of providing "direct interaction" for programming novices. Based on the above observations, it seems that the quest for direct interaction is not always necessary

### *D. Behind Learners' Questions*

When a learner raises a question, it may not mean that he needs immediate help. For example, a learner may propose a question because he wants to give the instructor a good impression and receive a higher score. On the other hand, the fact that no questions are raised in a class doesn't mean that everyone understands the content. For example, a learner having trouble with all of the content may not know what to ask.

Therefore, the over emphasis on the total number of questions raised may seriously distract learners' focusing.

### *E. Classify Questions Raised*

We may roughly classify learners' questions into two types: trivial and nontrivial. The trivial question is defined informally as those questions that can be answered quickly by a simple demonstration, simple example, or short explanation.

Frequent interactions put a heavy burden on the instructor and novices [6]. Better teaching/learning quality and grading policy will dramatically reduce the number of trivial questions. In programming language course, we ask:

*"For novices, is direct interaction more important than imitation and repeated practices?"*

### *F. ALPT: Active Learning and Passive Teaching*

In active learning, instructors reserve some course content and pose some questions to their students. Through self-exploration or teams, students answer the questions and understand the course content. The whole learning process involves students' active learning and frequent interactions. We characterize this kind of teaching strategy as ALPT ("Active Learning and Passive Teaching").

The ALPT strategy is still widely adopted. Unfortunately, one feature of ALPT strategy is that it often takes much longer than expected. Therefore, in a time of knowledge explosion, we may not be able to train enough qualified person in time using this kind of strategy.

### *G. BCL and ATPL Concepts*

Skilled people and unskilled people look at things differently. A trivial thing for a skilled person may be a big theory for an unskilled person. *In a time of knowledge explosion, if one does not have enough solid background on some specific fields, then "self-investigation" is synonymous with "waste of precious time."*

While dealing with the knowledge-explosion problem, improving "learning efficiency" is a more constructive strategy. We propose the following BCL (Baby-Centered Learning) concept:

*When facing a new course (topics), let the instructor "thoroughly" study it first. After fully absorbing the course content, the instructor will try to transfer the content to learners in the* ***shortest time possible****.*

We characterize this new teaching strategy as ATPL ("Active Teaching and Passive Learning"). This strategy emphasizes the instructor's (thorough) demonstration and explanation and the learner's repeated imitation. This is very different from the ALPT strategy that emphasizes the learner's self-exploration.

*"Programming learning" is a special kind of "language learning" that requires 100% of correctness. Therefore, any premature question will take up lots of time. We do not neglect the fact that active learning is very important. Instead, we only want to emphasize the fact that in order to learn faster, learners should be equipped with sufficient background prior to starting active learning*

### *H. Active Teaching Leads to Huge Data Size*

The adoption of the ATPL strategy in traditional education puts a rather heavy load on the instructor, especially for a large class in which the students' backgrounds differ widely. Thorough demonstration implies large number of images and clear explanation implies lengthy audio data. Therefore, the adoption of the ATPL strategy leads to a huge volume of course content In Table 1, some multimedia data sizes of an implementation of C language using the ATPL strategy are provided [5]. Putting such content (Table 1) on a web site would result in lengthy download for even just one topic. Therefore, it is not a good idea.

### *I. Teaching/Learning Cost*

Teaching and learning cost is a crucial issue; yet, it is often omitted in academic discussions. Distance learning is very expensive. Before a distance learning environment can be provided, costly infrastructure needs to be established. In addition, better local equipment is often required in order to get a better connection to the Internet. In distance learning, the total cost for a learner is the cost charged by the content server and the cost of connecting to the web (e.g., telephone bill). For those in poor countries, the expensive infrastructure makes distance learning a dream.

Therefore, in designing a new education model, the factor of cost should be considered seriously

## *5. SEARCHING FOR A NEW EDUCATION MODEL*

### *A. Proposing the CD-Type Learning Model*

We need a new learning model with the following features:

1. Lower $T_{ccd}$ value.
2. Lower $T_{mli}$ value, say, equal to 0, if possible.
3. Good course content.
4. Low learning cost.

Based on these four features, a new education model, named "CD-type learning model," is proposed here. It is defined as a software system that satisfies the following conditions:

1. Contains complete pre-recorded and at least one semester's course content.
2. Course content is well ordered and organized in units, and can be displayed in any order.
3. Course content may not contain video; however, it should contain clear images and audio.
4. Everything is stored on CDs and can be executed without using any browser.
5. Provides the function of replaying.

Condition (1) guarantees that the $T_{ccd}$ value is low. Condition (2) will result in a smooth learning curve.

In the traditional class meeting, learners often look attentively at the blackboard and listen carefully to what the instructor says. If there were nothing on the board, then they would watch the instructor. Therefore, in the class meeting, the content (could be transparencies) and the audio explanations are the most important elements. Therefore, condition (3) would minimize the differences between a real class meeting and a software education environment. Clear images and audio will help novices' understanding.

Condition (4) makes sure that the learning process isn't interfered with due to content delay, and also provides a portable learning environment with affordable cost. The differences between pre-recorded content on CDs and downloadable content from a web site are the learners' affordability (cost

of the infrastructure) and the time needed to access the content (complete the whole downloading). Condition (5) provides a self-pace learning environment.

Since the CD-type learning model contains complete course content, the $T_{ccd}$ value depends on the instructor's lecture speed. In general, it is less than 48 hours for a 3-credit hours' course. Moreover, since the $T_{mli} = 0$, the $T_{cm}$ value depends on the quality of course content and the individual learning pace.

*One thing we would like to point out here is that the name "CD-type" was coined at the very last stage of the implementation of this new model. The adoption of the CDs simplifies everything and provides better performance.*

The major differences between the CD-type model and the conventional used CAI (Computer Assisted Instruction) system are that CAI systems typically only focus on some particular topics and are used as a supplementary learning tool. Whereby, the CD-type model is designed as an alternative to current learning models.

### *B. Implementation*

The proposed CD-type learning model focuses on the issues of high learning efficiency and low learning cost. MPTE system (Multimedia Programming Training Environment) is one of possible implementations based on CD-type model and the ATPL strategy. MPTE integrates the roles played by the instructor, textbooks, and the computer. It provides the following functions:

1. A learning tree is used to give learners an overall view of the course content.
2. Each node of the tree corresponds to a special topic and is associated with a multimedia file, or a source program file together with a comment file.
3. The multimedia file corresponds to instructor's lecture and can be replayed at any time.
4. MPTE is a learning platform. We may link it with a designated compiler or interpreter. If we do so, then MPTE becomes a particular learning environment for that language.
5. Learners can do more practices within this environment by modifying demo examples and performing more drill problems provided in the system.
6. The node constructed by the course content designer is protected and cannot be modified.

MPTE provides a very fast learning cycle and the cost to a learner is just the price of those CDs'. Therefore, even a low-end PC system will provide an excellent learning environment.

### *C. Current Status of the MPTE System*

Currently, the MPTE system serves as a programming course learning platform running under Windows 95/98/ME systems. The whole system includes the MPTE system and a set of content designing tools. We used Borland Builder C++ to write the source code. It is about 22,000 lines of code. The complete course content is shown Table 2.

## *6. CONCLUSIONS*

Using students' questionnaires/scores to judge education models is too subjective. We need some metric systems to judge the superiority of different education models. Learning efficiency is one criterion to judge the superiority. In this paper, we transform abstract concept of "learning efficiency" to quantifiable parameters. These parameter values reveal some properties of education models. In distance learning models, parameters defined in this paper are not adequate to tell us about their superiority. Yet, measured values lead us to identify learning bottlenecks of the traditional education model, and we are very surprised to notice that the learning inefficiency is in fact an intrinsic property of the traditional education model itself. Parameter values also provided us guidelines for developing better education models.

Constant failures diminish ones' learning momentum very quickly. Therefore, we suggest the strategy of ATPL for novices. *These results might completely contradict our old beliefs. We emphasize the fact that*

*learners will learn better if they are equipped with sufficient background and skills before starting active, interactive learning .It is up to the instructor to decide when to start the traditional ALPT strategy.* ATPL strategy contradicts conventional pedagogy theories about active learning, and the implementation of the ATPL strategy will result in a huge data size for course content.

All of these results lead to the proposal of the CD-type learning model. Based on this new model, a possible implementation system called MPTE was designed, implemented and tested. MPTE provides a very efficient programming-learning environment. Compared to the distance learning, the CD-type model can save the high cost of the infrastructure as well as decrease the learners' learning cost and time. How to design a good course is an important issue; however, it is beyond the scope of this paper. "Good" course content take time to design, test, and modify. As soon as the content converges, novices will be able to learn programming languages very fast. Until then, programming will no longer be a privileged learning activity.

### *ACKNOWLEDGMENT*

I would like to thank anonymous referees for valuable suggestions.

### *REFERENCES*

[1] Griffin, R., "Management", Houghton Mifflin, Internet-Based Education, 4th edition, 1992.
[2] Tsichritzis, D., "Reengineering the University", CACM, June 1999, p93-100.
[3] Macek, T., Manova, B., Kolar, J., "Global Cooperation Project in Computer Programming Course", SIGCSE'99 Proceedings, p208-211.…
[4] Carrasquel, J., "Teaching CS1 on-line: the good, the bad, and the ugly", Proceedings of SIGCSE'99, p212-216.
[5] Shyu, Y., Chen, P., "MPTE: Multimedia Programming Training Environment", Proceedings of ICMCS'99, Volume I, p822-826.…
[6] Doube, W., "Distance Teaching Workloads", SIGCSE'00 Proceedings, p347-351.
[7] Anderson, R., Dickey, M., Perkins, H., "Experiences with Tutored Video Instruction for Introductory Programming Courses", SIGCSE'01 Proceedings, p347-351.
[8] Rapanotti, L., Blake, C., and Griffiths, R., "eTutorials with Voice Groupware: real-time conferencing to support computing students at a distance", ITICSE 2002 Proceedings, p116-120.
[9] Ari Korhonen, A., Malmi, L., Myllyselkä, P., Scheinin, P., "Does it Make a Difference if Students Exercise on the Web or in the Classroom", ITICSE 2002 Proceedings, p121-124.
[10] Zupancic, B., Horz, H., "Lecture Recording and Its Use in a Traditional University Course", ITICSE 2002 Proceedings, p24-28.
[11] Mcdonald, M., Dorn, B., Mcdonald, G., "A Statistical Analysis of Student Performance in Online Computer Science Courses", SIGCSE'04, p71-74

| Topics | Data size (Bytes) |
|---|---|
| Turbo C++ | 57.6 M |
| Borland C++ | 36.9 M |
| Call by Value | 120 M |
| Recursive Function | 89.9 M |
| Linear Sort | 71.6 M |
| Bubble Sort | 142 M |
| Recording Format: 22.05 KHz, 8 bits stereo | |

Table 1. Media sizes of some course topics

| Title | Disc no. | Size | Lecture length |
|---|---|---|---|
| BASIC | 3 discs | 1550 MB | 393 min |
| C-Stage 1 | 3 discs | 1514 MB | 408 min |
| C-Stage 2 | 3 discs | 1650 MB | 461 min |
| C-Stage 3 | 3 discs | 1600 MB | 450 min |
| Content of C language is divided into 3 stages. | | | |

Table 2. Sizes of designated course contents

**Yuh-Huei Shyu** is an associate professor in the Department of Information Engineering, Tamkang University, Taiwan, R.O.C. He has a Ph.D. in Mathematics from Illinois Institute of Technology and a M.S. in Computer Science from University of Arizona. His current interest is in knowledge transfer, especially in the techniques of transferring knowledge in multiple times of learning speed.