eラーニング・ファンダメンタル
eLearning Fundamental

Chapter 3 Evaluation techniques in e-Learning

Learning Objectives:

Be able to explain, using some examples, what Kirkpatrick's four steps are.
Be able to explain, using some examples, what the three elements to clarify objectives are.
Be able to explain how to use three tests to clarify the entrance and the exit.
Be able to explain five categories of learning outcomes and the evaluation method for each.
Be able to explain the purpose of formative evaluation and its techniques.

Summary of this Chapter

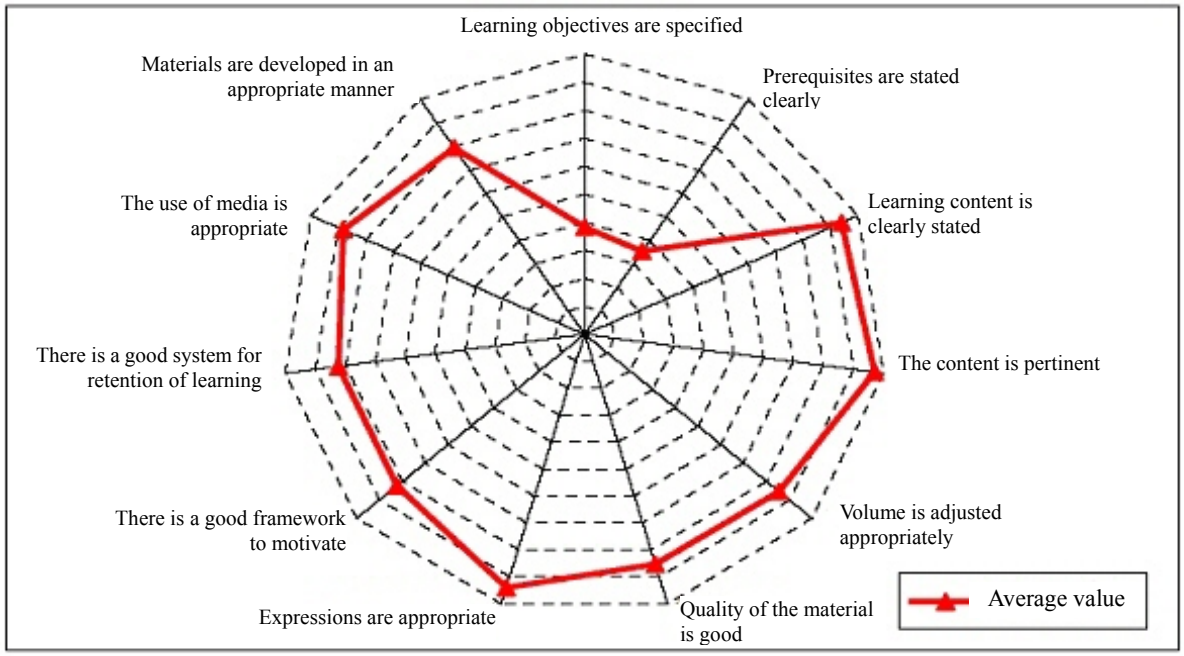
- The results of research on 17 cases of typical IT-related WBT courses showed that the scores for such items as "learning objectives are specified" and "prerequisites are clearly stated" are extremely low. The most important evaluation technique is to clearly state the objectives to the participants of the course.
- Kirkpatrick's four levels are Reaction, Learning, Behavior, and Results. They have been used for more than 40 years as a standard evaluation scheme for corporate training.
- The three elements of clarification of objectives are behavior verb, test conditions, and passing criterion. Unless you clarify the objectives, you cannot check if you have reached them, and you cannot be sure that the objectives are appropriate.
- The tests used to clarify the entrance and the exit are pretest, posttest, and entry test. By combining these three, you can clarify what you teach and to whom you teach by specifying necessity, effectiveness, and qualification, respectively.
- As frameworks for classifying learning objectives, the Bloom's Taxonomy and Gagné's five learning outcomes are widely known. The five outcomes are verbal information, intellectual skills, cognitive strategies, motor skills, and attitudes. The tasks are classified according to the characteristics of the learning objectives that are not only different in terms of the way of measuring the degree of achievement, but also in terms of the best way to support the learner to achieve them.
- Formative evaluation is a process to collect trial data before the material is completed with an intention to improve the quality of the material. As formative evaluation is particularly important for the development of self-learning materials, a variety of techniques have been accumulated.

Section 1 Merits and demerits of making learning objectives known

One of the studies which analyzed the evaluation perspective is "Study on promotion of distance learning system" conducted by CAIT and funded by The Mechanical Social Systems Foundation (The Mechanical Social Systems Foundation, 2001: they also published a 2002 research report that also contains reports on the result of comparative study of 2001 research and 2002 research). In this research, they selected 17 cases of typical IT-related WBT courses, and researchers who were familiar with the contents evaluated them based on the prescribed study items. In relation to the learning content and evaluation, they obtained the results shown in Figure 3-1. From the result, we can see that the scores for such items as "learning objectives are specified" and "prerequisites are clearly stated" are extremely low.

This study does not reflect any learning objectives or prerequisites that were stated on the WBT site as a reference for course selection, because the object of the study is only those statements that appear after the commencement of the course. They explain that this was one of the reasons why the scores were low. However, even if we take this into account, we still cannot say that the scores were good. It is also worth noting that, according to the study report, "Those courses that were first developed overseas then translated into Japanese tend to state learning objectives and/or prerequisites more clearly (p. 17)."

The first step of evaluation is to state the learning objectives. In this chapter, a variety of evaluation techniques will be introduced.



Source: The Mechanical Social Systems Foundation (2001), p. 17.

Figure 3-1: Results of evaluation of the learning content of IT-related WBT courses

Section 2 Kirkpatrick's four levels: A 40-year established scheme of corporate training evaluation (written by Hidekuni Komatsu)

3-2-1: Why do people need methods for educational evaluation?

In an information society, business becomes more and more software and solution oriented. In addition, with the ever-faster pace of change in the business environment, customer needs, and technology, education is becoming increasingly important. However, although a large amount of money has been invested in employee training in Japan, the effectiveness of the investment is not evident. As a result, whenever the economy faces a downturn and the performance of the company suffers, management targets training for cost reduction, resulting in contraction of training activities. This is because of the culture that considers training a cost, which stems from the lack of clarity on to what extent employees are satisfied with the training or how the training is contributing to business performance.

This could become a factor to misguide the future of the company, unless those involved in corporate training make sure that the human resource development (HRD) sector would not easily allow cost cutting. It should be done by obtaining proper awareness on the part of top management, by making their utmost efforts to demonstrate how training really benefits the business, which they do by carrying out evaluation of training. In addition, you are required to improve the management and system of training by carrying out evaluation of training to turn the Plan-Do-Check-Action (PDCA) cycle of training.

Figure 3-2 is a conceptual diagram of the objectives of educational evaluation. In relation to training, there is feedback to the planning/development/delivery stage of the training as well as feedback to the developmental stage of the training strategy. Sometimes the former is referred to as formative evaluation and the latter summative evaluation.

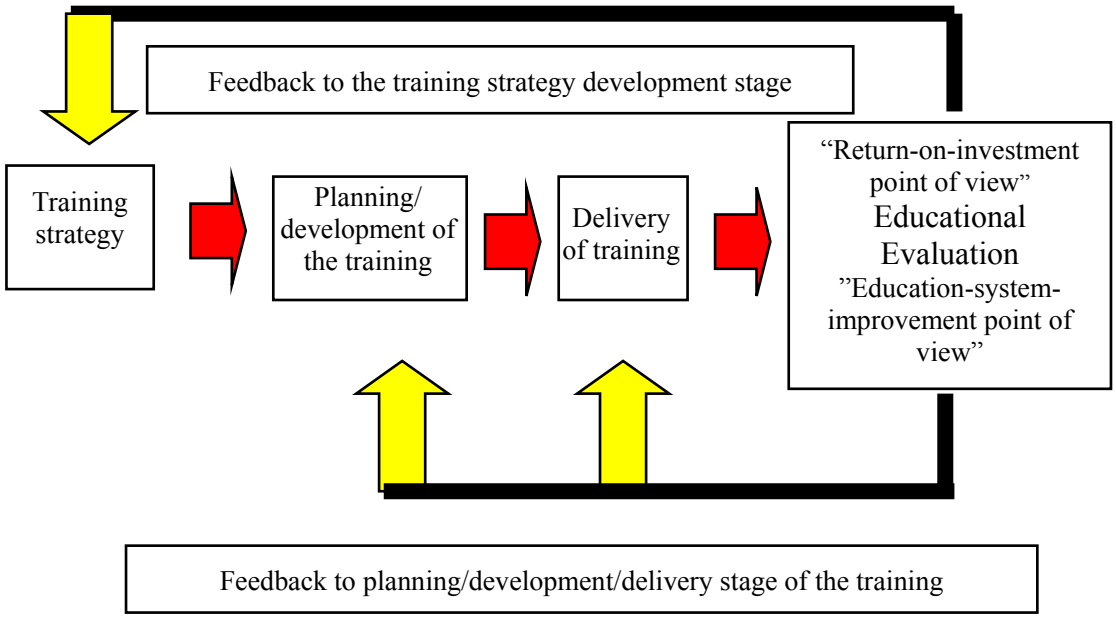


Figure 3-2: Objectives of educational evaluation in corporate training

The viewpoints of educational evaluation include the "return-on-investment point of view" and the "education-system-improvement point of view." The "education-system-improvement point of view" is such evaluation that is used to improve educational system and/or programs and is concerned with collection of data in relation to curriculum, material, instructor, and learning environment to identify such areas that should be improved or strengthened. The "return-on-investment point of view" is concerned with judgment as to whether continuation of the program in question is justified from the business management point of view, measuring the return on investment; i.e., to what extent educational objectives have been achieved, in order to make the business decision of whether or not to continue the investment (i.e., training).

3-2-2: Kirkpatrick's four-level evaluation model

Figure 3-3 shows Kirkpatrick's four-level evaluation model that carries out the course evaluation in four levels: Reaction, Learning, Behavior, and Results.

Figure 3-3: Kirkpatrick's four-level evaluation model

Level 1: Reaction	Satisfaction
Level 2: Learning	Understanding
Level 3: Behavior	Utilization in practice
Level 4: Results	Contribution to performance

Kirkpatrick, an American business scholar, proposed this model in 1959. Even now, after more than 40 years, Kirkpatrick's evaluation model is still being used unchanged as a commonly shared concept. Kirkpatrick even now never fails to deliver a lecture at ASTD every year and on other occasions, attracting large audiences. The familiarity of this evaluation method is such that it has now become common scheme shared by all. Although Kirkpatrick's evaluation is often expressed as "levels," sometimes it is expressed as "steps," indicating procedural order of "the steps to go from level 1 to level 4."

Summarizing the trend of the evaluation/training in the US that lies in the background of Kirkpatrick's evaluation model, it is characterized by the followings: (1) Awareness that the participants are customers, (2) Provision of such training that can satisfy the strategy, objectives, and policy of the business, (3) Competition between the in-house training department and external training providers, (4) Emphasis on the educational evaluation and efforts to improve the content of the training to enhance the degree of customer satisfaction, (5) Computerization of the evaluation system to speed up the evaluation process, and (6) Shared notion that evaluation has to be done.

In instructional design, everything is based on the notion that the participants are customers. There is no such notion that the provider of the education is a superior being who belongs to another world. As the purpose of training is to provide such courses that can satisfy the strategy, objectives, and policy of the business, evaluation is carried out for that purpose and used as a means of making improvement. The in-house training section is run by professionals such as instructional designers, and they have the notion that they are competing with external training providers. For that reason, they put emphasis on training evaluation and make efforts to improve the content of the training to enhance customer satisfaction.

In carrying out the evaluation, they have to communicate with learners to obtain the data. To achieve that, they have computerized the evaluation system and are carrying out the evaluation speedily. In Japan, evaluation is not readily acceptable, because when you talk about the evaluation of education, people tend to imagine that it means evaluation of the person him/herself or the ability of the person as a whole. However, in the US, based on the notion that training is an investment, people have no problem accepting the idea that evaluation must be done.

Looking at the rate of implementation of educational evaluation in the US, Level 1 evaluation is implemented in 92-95% of the cases. However, the rate of implementation is still not very high for other levels, with Level 2 evaluation being at 30-34%, Level 3 evaluation at 11-14%, and Level 4 evaluation at only 2-3% (source: ASTD, proportion by companies). See Figure 3-4 for the relationship between the levels of evaluation and the degree of difficulty/effectiveness of the training. It is said that Level 1 and 2 evaluations are carried out immediately after completion of the education, and Level 3 and 4 evaluations are carried out 3-6 months after the course. Whereas, even in the US, the rate of implementation is as low as reported above, I have never heard of any report of Level 3 or 4 evaluations being conducted in Japan.

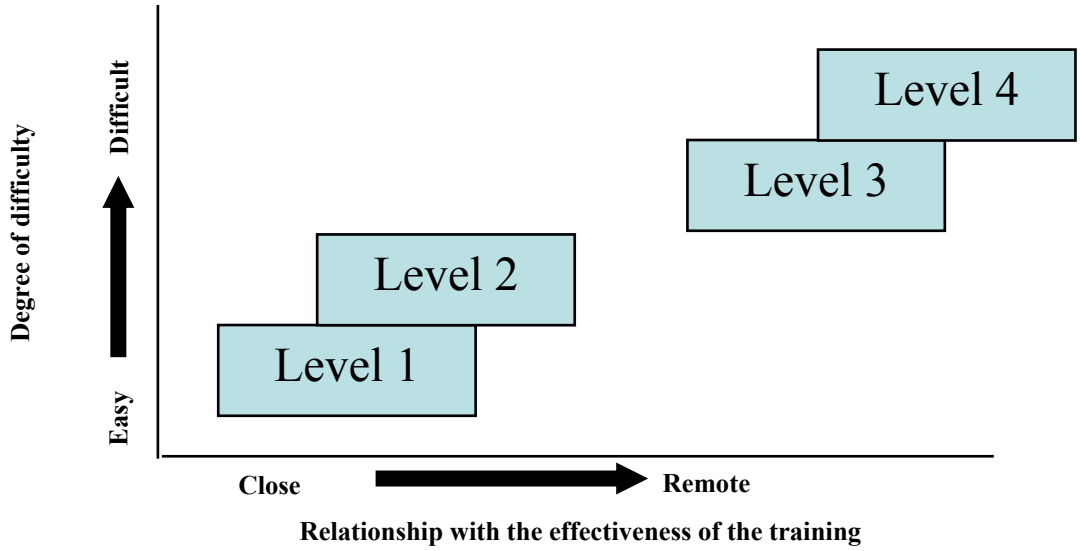


Figure 3-4: Relationship between levels of evaluation and degree of difficulty/effectiveness of the training

3-2-3: Examples of educational evaluation techniques.

■ Level 1 (Get hold of learners' response)

Forms of Level 1 survey to check the satisfaction level include questionnaires to participants/instructors/managers, collecting information through the complaint system, and computerized questionnaires. Basically the Q & A format shall be used; however, if you want to know the participants' impressions in concrete terms, ask them to put their impressions in writing. If you want a higher level of objectivity, use 3-10 multiple choice questions, and ask them to mark the answer so that you can quantify their responses. Ask them to write any voluntary comments to supplement the multiple choice questions.

If the course is divided into several sessions, it is desirable that you conduct the questionnaire immediately after each session. Alternatively, if there are organic links among sessions, you may conduct the evaluation for all the sessions at the same time at the end of the course while viewing the course as a whole.

■ Level 2 (Measure the level of knowledge/skills they learned from the course)

Types of Level 2 (understanding) evaluation methods include passing rate of the test, role-playing, checksheet of acquired skills, pre and post test, and comparison of skills before and after the course via e-mail. It is desirable that the test be conducted in as simple a manner as possible.

You may want to stress to the participants that the purpose of the test is to collect data for assessment and improvement of the course, rather than for assessment of the knowledge/skill of the participants. As for the content of the test, it is important for you to take advantage of the pilot course to improve its reliability.

■ Level 3 (Find out the utilization in practice)

Types of Level 3 (utilization in practice) evaluation methods include work observation by senior staff, observation of skills by third parties, questionnaires to/interviews with the participants/senior staff/junior staff, comparison of behaviors before and after the course, comparison of behaviors with a control group, and satisfaction survey of customers and employees via e-mail.

If it is assumed that, in measuring the return of investment (ROI) of training, what you have learned must be reflected in the way you carry out your actual work, Level 3 (behavior) evaluation is a very important point to check. As for the target of the survey, you should consider not only the participants but also other observers, including the senior staff. In addition, when you conduct an interview, depending on the situation, you might choose interviewers from external research organizations, in order to maintain neutrality of survey.

■ Level 4 (Looking for contribution to performance)

Types of Level 4 (contribution to performance) evaluation methods include comparison of work hours before and after the course, comparison of performance data before and after the course, comparison of income before and after the course (consulting firm), and comparison of income with a control group. In case of skill-related training, quantification of contribution is relatively easy. Research at this level is made difficult, because it is difficult to relate what was achieved by the training with its result.

One of the techniques of survey is a method in which the result is estimated based on the improvement made on an important step leading to the result. For example, it is known that, for sales representatives of life insurance, the telephone call to first-time customers to make an appointment to visit has an important influence over the success or failure of gaining a contract. If the rate of appointment making is doubled after training which teaches how to make an appointment on the phone, the productivity of the sales representatives is estimated to be doubled. Estimation of the performance can be made in this manner.

3-2-4: Notes on the introduction of educational evaluation

In the American environment, there is no uncomfortable feeling towards evaluation, or it is assumed that evaluation is always performed, because ID is used for the design/development of any training course. In contrast, you have to be aware of the environment in Japan, where people tend to identify the evaluation of education with evaluation of the person, resulting in an attitude of not accepting the result of evaluation in a straightforward manner because they consider educational evaluation to be difficult.

There are two reasons why evaluation of education is not carried out in Japan. One of them is that in Japan there has been a culture in which the evaluation of education is considered almost equal to the evaluation of the person as a whole. One thing that stands in the way of educational evaluation is that they have been carrying out practices such as gathering cadets together in a training course for the sake of screening. In addition, there is a tradition that people on the floor tend not to have full trust in the HRD section. As a result, even if they offer a course which is directly linked to the work, people tend to accept the result with prejudice. To break away from this kind of culture, for one thing, our only course of action is to keep achieving a track record of implementing training that contributes to development of ability that is directly linked to the work people do, to perform one's job so that we can gradually but steadily win the trust of the people in the company.

Although, in theory, educational evaluation should be introduced together with the introduction of the new education system, considering the cultural history mentioned above in Japan, it would be practical for us to propose, first of all, a style of training which contributes to work, establish a good track record, and then introduce evaluation methods little by little. It is not difficult at all for us to imagine a situation where, even if we practice such a good thing as instructional design, as people are not sufficiently open-minded to recognize its benefit, we would encounter resistance on the basis of misunderstanding.

For us to introduce educational evaluation smoothly, people have to share the notion that training is only useful if it contributes to the improvement of practicable ability to perform one's job. We have to use a systems approach for design/development/delivery of training so that we can build a framework in which we can positively make use of evaluation results. After that it seems that a bit of lead time is needed, because it would take some time for the evaluation method to take root.

Column: Philosophy of educational evaluation seen at AT&T

Let us take the example of AT&T, to look at the philosophy of educational evaluation. As for the distribution of responsibility for HRD, they have developed a system where responsibility is shared by the training department and business divisions, each of them playing a specific role. In terms of philosophy of training, the border between training and the actual work has been diminished. Training is valued so long as it has positively influenced the business. It is thought that the purpose of educational evaluation is to collect data concerning Levels 2, 3, and 4 to confirm the degree of influence that education has exerted over the business.

In a company which has advanced corporate training, in many cases each business division on the floor includes an HRD section, and such communication methods as holding cross-sectional meetings of the HRD sections of all divisions are used to set the orientation of the company as a whole or to unify training techniques. Once the size of the company has reached a certain level, for the sake of making training contribute to the business, it is a good idea, in terms of setting educational objectives, for the training assets to be located in the business divisions. On the other hand, it might be possible for giant enterprises, etc. to let one department manage corporate training in the form of human capital management (HCM), provided they are capable of offering individual-centered training, provision of information, and competency management functions at the same time to the whole company together.

Evaluation of corporate education, together with competency management, would be one of the most difficult frameworks to implement in Japan. In the advent of an information society, needs of the users, technology, and the contents of service keep changing rapidly. It is an absolute must for the companies in the future to put in place a learning system which can deal with such changes. Many people do not exercise and keep eating too much even though they are in danger of suffering from diabetes. Similarly, only a handful of companies would take up the challenge to optimize themselves for the times. Nevertheless, the changes are certainly coming, gradually but rapidly. One thing is certain: those companies and schools that are ready for the change will survive.

(Section 2 Written by: Hidekuni Komatsu)

Section 3 Three elements to clarify objectives: behavior term/test condition/ passing criterion

The Three elements to clarify objectives are behavior term, test condition, and passing criterion (Suzuki, 2002: See Chapter 3; exercises are available on the Web site). As it is said that the most important thing in the systems approach is to state what the objectives are in such a manner that can be understood by anyone, so far a variety of ways to clearly statement of the objectives have been proposed. Not only are the objectives used to confirm the success of the Level 2 (Learning) of Kirkpatrick's evaluation, but they are also used for the purpose of having the learning be focused on the objectives by letting the learners be aware of them from the start of the course. (cf. Event 2 of Gagné's nine events of instruction)

Let us think about the test condition in particular. If you think that to study is to memorize, you think it is natural not to have any reference material at your side when you sit for a test. However, by accepting the idea "As long as you can do it, it's OK to have some reference material with you," you do not have to force unnecessary memorizing. This kind of thinking leads to the idea of job aid (or EPSS).

Which of the following two objectives do you think is stated more clearly? (Suzuki, 1987)

A. Understand the method of clarifying the learning objectives

B. Given the list of three points to clarify the learning objectives, be able to restate all the unknown objectives if they are not stated in accordance with the three points. Note that the three points are statement of the objectives in behavioral terms, statement of test conditions, and statement of passing criteria.

When compared with objective A, objective B is more articulate in terms of what you want to teach, and to what extent you want to teach. As a result, objective B conveys its intention to the reader better. In other words, you can say that B rather than A is a more clearly-stated objective. A good way to distinguish a clearly-stated objective and one that is not is to think about how you can confirm whether or not a learner has achieved the objective. When you want to check if the learner has achieved objective A, you cannot really ask him/her "Did you understand that?" While there are a variety of ways to check if the learner "understood," ideas might differ from person to person. On the other hand, in case of objective B, all you have to do is to give the learner a few unknown and unclear objectives to check if the learner can restate them into clearly-stated objectives. In other words, if you have clearly-stated objectives, the evaluation method can be worked out easily.

3-3-1: Stating the objectives in behavioral terms

The first point to clarify objectives is to state them in behavioral terms. As what I want the learners to learn from the example above is "the method of clarifying the teaching objectives," it is certain that I want them to have an understanding of that method. However, while "understanding" means that the state of the matter inside (of the brain) the learner has been changed in one way or another, it is NOT possible for us to observe that change from the outside. While such objectives as "understand", "know" or "notice" are statements that straightforwardly tell what you want them to learn; it is not clear how one is to check whether or not the teaching was successful.

What about objective B? Here, whether or not the learner understood the method of clarifying the objectives is concretely stated "in behavioral terms"; i.e. in the learner's behavior to practically apply the method to restate the objectives. The behavior of "restating" in this situation is referred to as a behavioral objective. Needless to say, what you want the learner to learn in objective B is NOT the action of "restating" itself, because even if the learner comes to have a mechanical ability to restate without knowing why, that does not mean he/she has "understood" the meaning of "clarification of objectives." You have to be aware that the behavior of "restating" is used as an indicator to show the fact "understanding" in such a concrete way that is observable from the outside.

3-3-2: Stating the test conditions

The second point of clarifying objectives is to clearly state the conditions under which behavioral objective is assessed. In other words, the conditions of objective B are the part "given a list of three points" and the "unknown" part. With this objective, it is clear that the learner does not have to memorize the three points. All that is required is to know the meaning of each point, because at the time of assessment the list will be given as a part of the test question. In addition, you can see from the condition "unknown" that the assessment will not use any example that the learner has ever seen, such as a sample question or exercise. This is to prevent the learner from being able to give the correct answer by learning by rote what has been restated. This is because, to test the learner's ability to apply the skill, you have to check if the learner can apply it to a new example. Conditions in other cases may include things the learner can use or restrictions to be imposed, when he/she takes a test.

3-3-3: Stating passing criterion

As the third point of clarifying objectives, sometimes the criterion to judge whether or not the objective has been achieved is stated. In case of objective B, the part "all" corresponds to this. In this case, it is possible to replace it with, for example, "four out of five questions given to you." In some cases, the objective can include other criteria such as speed, as in "swim within one minute," or accuracy, as in "measure within an error margin of 5%."

For the purpose of clarifying objectives, it is not a bad idea to start with an objective such as A which states in a straightforward manner what you want them to learn. By applying the three points for clarification above, selecting behavioral objective and adding test conditions and passing criteria, you can gradually shape it into a clearly-stated objective. One of the methods that are effective in helping you to carry out this process is to make test questions to evaluate the performance of the lesson. It is worth trying to work out the clearly-stated learning objectives reversely from the concrete method of evaluation by creating test questions.

Column: Clarification of objective is required from an economics point of view as well

Mr Koshio, the author of "Economics of education," takes up the issue of what should be considered educational achievement from the standpoint of economics. Let me quote, among other statements I am interested in, a few of those arguments that objectives should be clarified.

> They are proposing such educational objectives as "improve basic scholastic ability," "nurture children who are rich in spirit," and "zest for living."

> However, this is hardly a system to properly evaluate to what extent such objectives are achieved. (Koshio, 2003, pp. 107-108)
>
> I come to wonder how they verify/measure the degree of success education has achieved in making the children learn "zest for living." As people, these days, have come to actively stress the importance of policy evaluation, if the government makes it one of its objectives to let the children learn "zest for living," the most important point in evaluating the education administration should be how successfully that objective was achieved. How in the world are they going to do this evaluation? (p. 109)

Pointing out that, "If the objective is too lofty, no one really cares about it. The more concrete and measurable the results of educational objectives are, the more desirable."(p. 110), he welcomes the move (quoted from "Yomiuri education email": an example from Setagaya Ward) toward indicating it in concrete figures. For example,

> "Deep-thinking children"→ "Double the number of books children can borrow from the library"
> "Make them always have 10 or more kinds of flowers in bloom from April to October"
> "Make the percentage of students who are assessed level 4 or above in 5-level assessment more than 90% in terms of the satisfaction rate in relation to the pathway they selected at graduation."

> The author does not think that all the educational achievements should or could be evaluated in terms of quantities. For example: the words of your beloved primary school teacher or the bitter-sweet memory of the fight you had. However, is it really desirable for us to consider the educational achievement utterly unmeasurable to the extent that we refuse any external evaluation? In my opinion, education must not be a sanctuary free from any criticism. There must be something in education that can be verified in a proper way. If you refuse even that, it cannot be helped that people criticize you as being self-righteous (p. 111).
>
> There are two reasons why we verify the effectiveness of education from the economics point of view. (1) There are costs (for example, fees and taxes) involved in education. It is also the most important area of concern for the study of economics whether the educational institutions are successfully producing such results that correspond to the costs, because one of the main areas of concern for economics is how effectively limited resources can be distributed. (2) For the sake of fairness, we, economists, are also concerned with how differently the achievement of education is reflected in each child. Especially in the case of compulsory education, the wide range of differences casts doubt about the way the government is involved in education.

Source: Takashi Koshio (2003), "Economics of education", Nihon Hyoron Co. [In Japanese]

Section 4 Three tests used to clarify the entrance and exit

The tests used to clarify the entrance and the exit are pretest, posttest, and entry test. By combining these three, you clarify what you teach and to whom you teach so that you can clarify necessity/effectiveness/qualification, respectively (c.f., Suzuki, 2002, Chapter 3).

Clarifying learning objective means clarifying the "exit." In other words, the learners come to know "what is going to be taught" and "what they are going to learn." The judgment of this is made by a posttest. In other words, the posttest is a test to judge whether the learner has achieved the learning objective. If the result is a failure, then the learner takes the posttest again after relearning the course.

If before taking the course the learner knows at least some content of the course or might be able to do something included in the content, a pretest and entry test shall be implemented before the course starts. Pretest is the same exit level evaluation as the posttest, so if the learner passes it, there is no need to take the course and he/she should skip this course and move on to the next. If the learner fails the pretest, there is a need to take the course as expected and he/she starts the course.

On the other hand, the entry (readiness) test is an assessment to check if the learner has already learned the basic skills which will not be dealt with in the course. If the learner passes this, he/she is "qualified" to start the course, and if not, he/she is judged as "NOT qualified yet." In this case, the learner would be asked to (re)take a preliminary course before trying to start this course again.

It is possible to combine the pretest and entry test to one test, when they are implemented. In that case, the test is used as a "diagnostic test" to determine from which level the learner should start. These two tests are used to clarify the entrance, determining who should learn, and to whom you teach. These tests play the role of a gatekeeper who shuts out those learners whose level is either too low or too high.

In group instruction, difference in readiness level becomes an issue. There will be those who have difficulty in keeping up with the pace of the course and those who get bored because the course is too easy for them. Whereas there are some people who think that this is a necessary evil, there are other people who think that if you do not deal with the individual differences in one way or another, you are not a good instructor. However, if you control the entrance and exit in a proper manner, it is possible to line up the level in terms of knowledge/skill of the participants. By properly analyzing the learning content, dividing it into a number of small units and controlling the entry/exit in a proper manner, you can make it possible not only to remove waste/excess/irregularity from the course, but also let everyone have productive time.

As people realized this in the US, they started to implement mastery learning. Although they say that there are merits and demerits, I think that mastery learning is a reasonable and really align to systematic approach of ID.

Column: TOTE model (Test-Operate-Test-Exit)

In the TOTE model, as shown in Figure 3-5, when you carry out a certain operation, you check, first of all, if the objective has been achieved. If the objective has been achieved, you will not carry out the operation, and will exit. If it was found that the objective has not been achieved, you carry out just a certain amount of operation, then check again to see if the objective has been achieved. If it has been achieved, you exit from the loop, and if it hasn't, you go back to the operation, repeating the cycle of check-operation-check. As shown above, the TOTE model is an abbreviation of Test-Operate-Test-Exit.

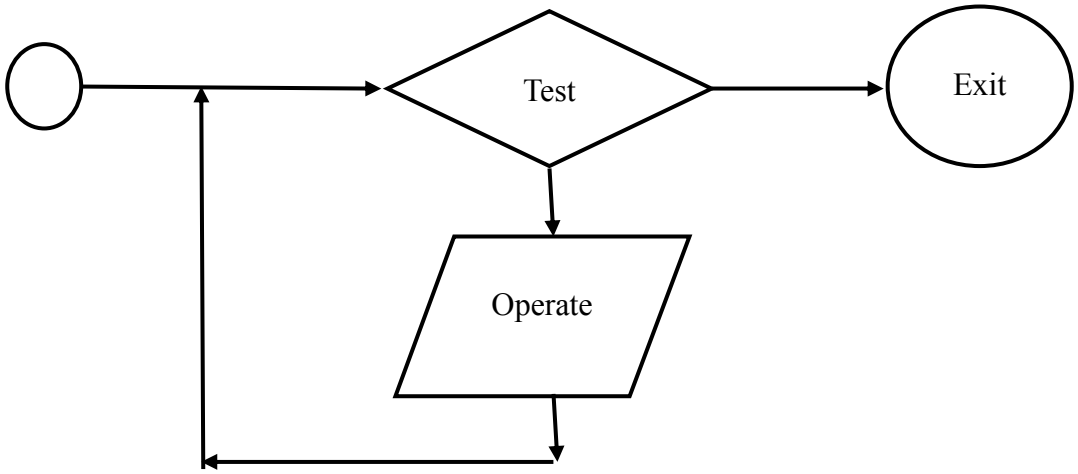


Figure 3-5: The TOTE model

The TOTE model is a diagram to show the way you make progress towards a certain goal, in that you keep checking to see if you have reached the goal. Temperature control of an air conditioner, for example, operates according to this model. While this is a very simple model, this model gives you the most important concept when you think about how to proceed with e-Learning.

This model is the basis for such ID concepts that you carry out the pretest to confirm if there is a need for the learner to learn it; only those learners who have a need to learn are required to take the course. Another idea from this model is that you carry out the posttest, and if the result is a fail, the course (operation) has to be repeated, whereas if it is a pass, the course finishes there (exit). This model gives you many useful suggestions, such as you do not just carry out operation, instead you start the operation only when it is necessary. By showing the learning objectives at the start of the course, it becomes possible for the learner to take such option as "I am already familiar with this, so I won't do this; instead, I will do other materials." Evaluation (test) is carried out not to give the participant just a score, but to see how the participant performed on the test to decide the next course of action.

Section 5 Nature of learning objectives and appropriate evaluation method: Bloom and Gagné

Various frameworks for classifying learning objectives have been proposed in the past with an aim to match up evaluation methods with the nature of the learning objectives. The most famous is Bloom's Taxonomy. Bloom's Taxonomy is a proposal to classify objectives into different categories as a framework for evaluation method. After Bloom's books were translated in to Japanese, it had a significant influence over those involved in education in Japan as well as those involved in education overseas (for details, see Kajita, 1992). One of the credible contributions was that it listed appropriate test questions to measure the achievement at high levels of cognitive ability. It challenged the contradiction in educational practices that even if the objectives were set at a high level, the actual test questions tended to be those that could be answered easily even by a learner who only has fragmented knowledge acquired by rote.

Bloom's taxonomy divided the educational objectives into the following three domains:

(1) Cognitive Domain: Objectives concerning reproduction of knowledge, development of intellectual skills, and so on

(2) Affective Domain: Objectives concerning change of interest/attitude/values and capacity to adapt

(3) Psychomotor Domain: Objectives concerning motor skills or skills of manipulation

In addition, each domain is divided into levels for the purpose of categorizing the objectives hierarchically from basic to higher levels. The cognitive domain is divided into six levels of objectives. They are, from the most basic level, (1) "Knowledge," information which you can recall, and use, if necessary, as what you have learned previously, (2) "Comprehension," the ability to understand the meaning of and use conveyed information, (3) "Application," the ability to apply acquired knowledge to a new task/scene or actual situation, (4) "Analysis," the ability to break down/reconstitute the problem to understand the overall structure of the problem, (5) "Synthesis," ability to combine elements together to create a new whole, and (6) "Evaluation," the ability to judge the value or meaning.

The affective domain is a domain concerned with such objectives as "how I think" or "how I feel." Depending on the degree of internalization of values, objectives are divided into five levels. They are, starting from the basic level, "Receiving," "Responding," "Evaluating," "Organizing," and "Characterizing." The psychomotor domain includes various motor skills which are dealt with in physical education programs and such skills as writing characters or manipulating typewriters. Bloom did not categorize this domain into levels. Although some people have attempted to categorize it, there is no agreed upon theory yet. Based on the above, people have actively tried to classify the objectives and categorize them into levels within each area of subject/content.

On the other hand, Gagné, famous as the founder of ID theory, classified the learned capabilities as shown in Figure 3-6 (Note: Exercise questions are available on the Web site).

Figure 3-6: Gagné's five categories of learned capabilities, and examples taken from study of foreign languages (Suzuki, 1989)

Category of learned capabilities	Details and examples taken from foreign language study
1) Verbal Information	To learn such facts or labels which, as a result of learning, you become able to "name." For example, to memorize the foreign word that corresponds to a concept you have already learned in your mother tongue (for example, "inu"; "inu" is Japanese for "dog"). You are considered to have learned the verbal information if you become able to state (either by saying or writing) "Inu means dog" when you are asked "What does inu mean in English?"
2) Intellectual Skills	Cognitive learning by which, as a result of learning, you become able to "apply" the learned capabilities to a new example which you have never encountered before. Intellectual skills are further classified into the following lower-order domains.
2a) Discrimination	Learning by which, as a result of learning, you become able to see (or hear) the "difference" between two things. For example, to be able to tell which one is different from others after listening to the pronunciation of *r*, *l*, and *r* in succession. In this example, you are considered to have learned the difference between the two; in other words, you became able to discriminate, if you can point out the one which is different, even if you cannot expressively tell which one is *r* and which one is *l*.
2b) Concrete & Defined Concepts	Learning by which, as a result of learning, you become able to classify those objects/events that belong to a particular group (concept) and those that do not. For example, to distinguish a grammar concept which does not exist in your mother tongue (for example, to distinguish nouns whose plurals have the same form as their singulars and those whose plurals do not) or to learn the deviation of concept between your mother tongue and foreign languages. Whereas when you learn the word "seat" as "verbal information" all you are expected to do is to describe it as "a thing on which you sit," to prove that you have learned that as a concept, you have to be able to correctly give an answer to such questions as "Is the chair in the movie theater a seat?" and "Is the bench in the park a seat?"
2c) Rules; Higher-order Rules	Learning by which, as a result of learning, you become able to apply rules. For example, you are considered to have learned the rule if you can apply, to a sentence, such grammar rules as the rule for converting a statement into a question and the rule for changing the verb conjugation in accordance with the subject or time. Further, an ability to apply a series of rules; i.e., to be able to make use of previously-learned rules to solve, for example, such a problem as conveying your intention to the other party in a certain situation, in particular, is called higher-order rules.
3) Cognitive Strategies	To learn how to learn. You are considered to have learned the cognitive strategies when you have become able to use, when necessary, those means that make the learning effective, which you obtained through your previous learning experience or learned from hints given to you from outside. One example is the method of memorizing words by using equivoque.
4) Attitudes	Such learning that helps development of an internal state which, when you encounter an opportunity to make an individual choice, affects such individual choice. For example, if you learn a positive attitude toward active communication with foreigners, you will take a course of action which would positively choose opportunities to communicate with foreigners. If you learn an attitude to speak English, you will spend more time on learning English than other things. In these cases, you are considered to have learned a positive attitude, as a basis for such actions as learning English or communicating with foreigners.
5) Motor Skills	Such learning which, as a result of learning, you become able to use those skills that involve (not just cognitive but also) muscular activity. Examples include pronunciation, which involves muscular activity of the mouth and the skill to manipulate a pen to write letters, which involves finger movement.

Note: Source: Suzuki (1989). The above table is a summary of the body of the text.

The five categories of learned capabilities are verbal information, intellectual skills, cognitive strategies, motor skills, and attitudes. These frameworks were worked out by classifying the tasks based on the nature of the learning objectives from the perspective of ID. These categories were to be not only different in terms of the way of measuring the degree of achievement, but also in terms of the best way to support the learner to achieve (see Chapter 7 below).

Learned capabilities in the cognitive domain are divided into three kinds. Intellectual skills mean to learn rules such as method of classification or method of calculation, as well as abilities to apply them to unknown examples (procedural knowledge). Verbal information means to learn abilities to restate such given information as names and names of an era (declarative knowledge). Cognitive strategies mean to learn abilities to make your own learning process more effective (learning skill). For intellectual skills, subcategories are defined based on the results of many years of study by Gagné.

Affective domain includes learning of attitudes. Attitudes, which is considered one of the learned capabilities, mean either positive or negative feelings, including "race discrimination" or "to learn mathematics," toward all kinds of objects, events, and situation. In addition, the motor skill domain includes such motor skills that you become able to achieve a certain task by moving your body (the whole body or a part of the body). Apart from those learning tasks that are dealt with in physical education, motor skills also encompass such as touch typing or pronunciation of a foreign language. See Figure 3-7 for a summary of these.

Figure 3-7: Gagné's five learned capabilities and clarification of entrance/exit

Learned capabilities	Verbal information	Intellectual skills	Cognitive strategies	Motor skills	Attitudes
Nature of the skill	Memorize specified items Declarative knowledge Reproductive learning	Ability to apply rules to unknown objects/events Procedural knowledge	Ability to make one's own learning process effective Learning skills	Muscular ability to move/control one's body	Emotion to choose/avoid a certain object, event, or situation
Action verb to show the classification of the learned capabilities	State	Distinguish Confirm Classify Demonstrate Form	Employ	Execute	Choose
Evaluation of the capability	Recognition or reproduction of previously presented information Either target all the items or use random sampling	Apply the rule to an unknown example: should not be reproduction of the rule itself Questions should target all types of tasks to check the applicable range	To be applied to the process of learning rather than the results. Use observation of the learning process, self-descriptive report, etc.	Demonstration: the knowledge of how to do it differs from the ability to actually do it Check if the wrist is used to do it accurately, speedily, and smoothly	Either observation of behavior or declaration of intention to take some actions Set up a situation. Deal with individual's choice rather than a general choice

Source: Suzuki, K. (1995), "Premier of instructional design from broadcasting", NHK, (extract of Table III-2)

Column: Developing the ability of self-evaluation and self-appealing by the use of a portfolio

Portfolio means a briefcase or a file. This is a word of foreign origin which has become the focus of attention in recent years as a means of assessment for overall learning activities in primary schools and junior high schools in Japan. Portfolio assessment is defined, for example, as "a method to store, in a file, compositions, reports, works, tests, and photographs or VTR media showing the activities that have been created by children/students in the course of their learning activities" (Glauert, 1999, p. 8 in translation).

It is generally thought that you do not store all the things you have created in the course of learning, because portfolio assessment is not a mere record but an assessment. In other words, it intends to develop metacognition to control one's own learning activities by (1) expressively telling the children what they have achieved, (2) letting them understand the reason why they are assessed high or low, (3) enhancing the feeling of fulfillment or self-esteem, and (4) showing what the next tasks are, through the process of selecting the items that are worth keeping and filing them in front of the children. This technique is similar to a security portfolio (a set of securities which is less affected by change by virtue of combining those securities that are not mutually linked to each other) in the sense that you have to decide what you should keep in the portfolio to demonstrate your learning performance the best way you can think of.

They say that in the UK students who are 16 years of old or above are obliged to prepare a portfolio by themselves as evidence to show that their learning performance is good enough to obtain a General National Vocational Qualification (Glauert, 1999). They are trying to educate their children in such way that the children can put together and arrange in a portfolio what they have learned and, on the basis of that, demonstrate their performance by expressing "This is what I have done." I can imagine that if we could train our children from junior high school level, in such a way that they can demonstrate their achievement based on the evidence they themselves prepared, we would have a very robust sort of children.

Moreover, Oda (1999) in introducing portfolio learning in the US, points out, "There are few schools in Japan which carry out comprehensive learning while clearly explaining the assessment point of view (p. 8)" According to Oda; in the US, they clearly explain the assessment point of view to the students before they start their learning. Clearly showing the assessment point of view in advance is considered to be essential for the portfolio learning because it is "the learning to realize the change that happened to them by themselves (p. 75)" Self-assessment and correction by feedback is an integral part of portfolio learning.

Take the case of "Market research"; in Minnesota, for example, students are obliged to submit a monitoring diary, graphs, and a short essay. They are shown assessment criteria for each submitted item; i.e., for the monitoring diary, whether it is easy to understand and whether it contains important points; and for graphs, whether the display is precise, whether patterns are found, and how patterns are shown. Students first of all check them by themselves assessing in three levels, "excellent, satisfactory, or to be improved" (this is called a "rubric"), and reviewing their learning activities while comparing their assessment with that of the teacher and develop their self-evaluation ability.

In Japan, it is often said, "In portfolio learning, not only are the results of learning assessed, but

so is the process." However, in my opinion, this is wrong. When you say that the process is assessed, it seems that there is a gesture of kindness such as "we take into account the effort of the children whose results were not very good." However, isn't that misguided kindness? When you take the amount of effort into account, should those children whose ability was so high that they did not have to make much effort to achieve the task be assessed low? Although portfolio could be used to check the process in detail, based on the discussion above, I feel that it is asking a severe question, "How would you demonstrate what you have achieved in the course of learning, based on what sort of evidence?" This question does not have anything to do with the process, but has to do with a series of achievements the children have made one on top of another (i.e., multiple products, rather than process). The evaluation by portfolio includes evaluation of self-evaluation capability (this is also one of learned capabilities rather than a process) which is concerned with the level of ability to, by skillful combination of one's work, demonstrate one's achievements in the course of the learning process. In terms of Gagné's taxonomy, it is possible to interpret it as that which aims to develop and evaluate the cognitive strategies.

In Japan, when it comes to assessment, it has been considered that while it is always the sort of thing that is handed down from the teacher, it is a very powerful thing, so powerful as to almost determine your value. With this kind of history as a background, people are still trying to work out a new form of evaluation. People might be just confused to suddenly hear such things as assessment to enhance yourself, assessment to know your own weaknesses so that you can work on them, or assessment to make you capable of demonstrating your own goodness. It would be far from easy for those adults who in their childhood were exposed to the conventional view of assessment to redefine the meaning of assessment. Nevertheless, it is another important responsibility of instructional designers to tell the people the role and significance of assessment in the learning process. With this awareness in mind, we must step up our efforts to make people aware that assessment is a powerful tool for the participants of the course when it is carried out in an appropriate manner.

References

Glauert (translated by Hideyuki Suzuki) (1999). "Portfolio assessment for teachers and children (for comprehensive learning/science)", Ronso Co. [In Japanese]

Katsumi Oda (1999). "Portfolio learning and assessment suitable for comprehensive learning", Gakuji shuppan [In Japanese]

Section 6 Formative evaluation of material to be used for e-Learning

Formative evaluation is an important step to positively carry out effective material development. Defined as "the process designers use to obtain data that can be used to revise their instruction to make it more efficient and effective." (Dick & Carey, 2001, p. 284-285), formative evaluation means the process by which, before the development of material is completed, learners try the material for experimental purposes to make improvement to the material (AECT, 1977, p. 265). The material that went through the cycle of formative evaluation and the revision is considered a material whose effectiveness has been enhanced based on the data obtained from actual learners. It is then going to be used as a product of the systematic process of ID. In other words, formative evaluation is a process of data collection to check such material that has been designed from a theoretical point of view based on the ID model. For example, selection and application of teaching strategy was made based on structural analysis of the learning task and/or theory of learning. In order to see if such decisions are valid, we conduct formative evaluation to check and make revisions, for the purpose of improving the material.

The term formative evaluation was used by Scriven in 1967 in an effort to clarify the educational evaluation by separating its role into two kinds of roles; i.e., summative role and formative role. The result of educational evaluation which has a summative role flows out of the agent of material development to be used to improve the way the material is used or awareness toward the material. On the other hand, the result of formative evaluation is used, as inside information, to improve the material which has not been completed. In other words, with formative evaluation, the object of the evaluation is neither the learner nor the "completed" material. You collect the data to improve the material as a step of the ID process. During these 20 years since introduction of the concept, researchers have been trying to systematize the methodology. The methodology of formative evaluation used in the current ID process model has been formed on the basis of these research efforts (for details, see Suzuki, 1987).

The techniques which constitute the core of formative evaluation are related to data collection from the learners. The data from the learners include not only those question items that are included in the pre/post test or materials for the purpose of identifying the stumbling points, but a questionnaire asking the learner to put forward his or her impression or proposals regarding the material, a record of learning time, etc. Other sources of information include the teacher who is in charge of the target students, experts of the subject matter, and professionals who specialize in instructional design. See Figure 3-8 for kinds of data obtained from the respective sources of information and the timing of collection (Suzuki, 1987). From these sources, you collect such information that is likely to be useful in revising the material, bearing in mind that the purpose of formative evaluation is to enhance the effectiveness of the material.

Figure 3-8: Kinds of data used for formative evaluation, source of information, and the timing of collection

	Source of information	Kind of data		
		Before trial	During trial	After trial
External volunteers	Instructor	• Appropriateness of the material	• Administration of material	• Opinion, impression, and proposal regarding the material
	Learner	• Result of entry behavior test	• Result of test incorporated in the material	• Result of posttest
		• Result of pretest	• Opinion regarding the material and its content	• Opinion, impression, and proposal regarding the material
		• Clearness of instructions and items in relation to pre/post test		• Elapsed time before completion
Within the team	SME	• Accuracy and freshness of the content and appropriateness of the vocabulary level and sample questions		
	ID specialist	• Appropriate application of learning instruction theory		• Appropriateness of the revision of the material
		• Appropriateness of the term		

Source: Table 1, Suzuki (1987). Note that some terms in the table are changed for the purpose of simplification.
Glossary: SME = Subject Matter Expert

■ One-to-one evaluation

The first stage of collection of data from the learners is called one-to-one formative evaluation. In this process, designer of the material guides the learners one at a time to go through the course of learning using the material, for the purpose of removing obvious errors from the material as well as the tests, and checking learner's response to the material. At this stage, it is important to create an atmosphere where the learners can feel free to view the material critically and express their opinion regarding any parts of the material with which the learner feels uncomfortable; for example, when the meaning is not clear. The designer must go through the learning process with the learner page by page, while trying to create such kind of atmosphere as mentioned above by, for example, interposing questions at appropriate places in accordance with a set plan.

It is proposed that one-to-one formative evaluation should include at least three learners, at the high/middle/low levels, in relation to the yardstick which is thought to be most influential to the success of the material. For example, if motivation is the factor which is most closely related to the effectiveness of the learning, pick from the target group the most motivated learner, one whose motivation is around the middle, and one whose motivation is not very high, to go

through the learning process with them one by one in the order mentioned above using the material. If they, the high/middle/low level volunteer testers, show different responses, work out some strategies such as creating branching-out or voluntary items. In the one-to-one stage, apply the same procedure not just to the material, but to other things such as tests and questionnaires as well to check and remove obvious errors before you go on to the next stage.

■ Small group evaluation

The second stage of formative evaluation is called small group formative evaluation. Whereas designer is involved in the learning process in the one-to-one stage, in the small group stage, you try to identify those problems that are associated with the case where learners go through the learning process independently. At the same time, you also check the areas which have been revised as a result of the one-to-one evaluation to see if they are effective. For the small group evaluation, you recruit 8-20 volunteer testers in order to simulate the target group as well as to examine the result in quantitative terms. For this reason, you have to pay particular attention to the way you select the volunteer testers, especially when the target group includes various types of learners. The results of the small group evaluation are classified in accordance with the learning objectives or characteristics of the learners to consider and implement revision of material. Although for small-scale material development the effectiveness would be sufficiently verified in many cases by conducting the evaluation at this stage, for large scale development, you go on to the next stage.

■ Field trial

The third stage of formative evaluation, called field trial, is an evaluation under such practical conditions as management of the material and relationship with other courses. Although you can carry out the field trial without one-to-one or small group evaluation, it is supposedly more effective to conduct one-to-one and small group evaluation step by step before you carry out the field trial. The first year of implementation in a sense includes an element of "field trial," because you gradually improve the quality for the second year implementation while you are actually implementing it and dealing with problems on the fly.

These processes of formative evaluation have many common aspects with the techniques used in the test operation of α and β version or usability testing recommended by systems technology. Although these processes are recommended in every area to ensure the quality of the final product, in the area of education, they tend to be neglected.

■ Reflecting the result of the evaluation to improvement

Three steps of formative evaluation are explained above. See Dick & Carey (2001) for details. In addition, taking the printed material as an example, a checklist to work out improvement ideas is proposed in Suzuki (2002), (note: you can also try it as a tool on the Web). Flagg (1989) introduces the practical example of formative evaluation in a project which utilizes a variety of media.

On the other hand, in many senses it is not clear how to interpret the data obtained from formative evaluation to relate them to the improvement of the material. It is said "In our approach to formative evaluation, we interpret the data in light of our instructional strategy and then make changes that seem to be indicated by the data and our understanding of the learning process." (Dick & Carey, 2001, p. 323)

See Figure 3-9 for Debert's improvement list as a hint to visualize what sorts of concrete actions are involved in the improvement of materials. Although this list includes 26 ways of improvement, equal in number to the letters in the alphabet, they are not comprehensive. Although the expression material improvement sounds simple enough, it is surprising to find there are such wide a variety of ways to deal with improvement of materials. In addition, Nathenson & Henderson (1980) argue that it is convenient to classify it into such four categories as Add (to add improvement to the material), Delete, Move, and Modify.

Figure 3-9: List of strategies to improve the material (by Debert)

Add	A. Add explanation of entry skills and knowledge. B. Add training to teach learners how to use the material. C. Add training for instructors who give the instruction using the material. D. Add an advance notice (advance organizer). E. Add illustrations. F. Add operational aid. G. Add examples. H. Add activities. I. Add feedbacks. J. Add exercises for transition. K. Add test items. L. Add motivators. M. Add diversity.
Simplify	N. Lower the level of complexity. O. Simplify the vocabulary. P. Use smaller units. Q. Adjust the material to larger units. R. Change the order. S. Delete information which is not very relevant. T. Delete activity which is not very relevant.
Others	U. Replace examples with those that are more relevant. V. Change the instructional media. W. Change the format of the material. X. Change the learners who were used in the formative evaluation. Y. Throw away the project. Z. No change.

Source: Suzuki (1987), Table 2 is quoted here while changing some terms.

Dick & Carey (2001) point out that you should not start modifying the material itself straight away even if the result of the formative evaluation was far from satisfactory. According to them, before you consider the modification of the material itself, you have to first of all recheck the test itself, then re-examine learning task analysis, followed by checking instructional strategy to see if they are mutually congruent. A check is also needed to see if communication with the learners was conducted properly. In some cases, posttest may include an inappropriate item, even if the material itself is appropriate, then the obtained data cannot be as good as you would expect. In addition, they also argue that, it is sometimes necessary to determine the priority of revision by considering whether the benefits of modifying the material justify the required costs.

Column: Can't the evaluation by evaluation agent (experts) be relied upon?

An experimental study reports that there was a deviation among those ratings that were given by 30 volunteers who compared 6 WBT packages. The report argues that we must obtain empirical data by way of formative evaluation rather than depend on the evaluation by experts, because the ratings above were completely opposite the ratings given by the experts (www.lguide.com). Can't we rely on the evaluation by the evaluation agent? The following is the summary of the study:

Objects of comparison: Six commercially available applications to learn Microsoft Access (price range: $20-$99). Among ten commercially available applications that they found, one was excluded because it was offered as CD-ROM, not WBT; another one was too expensive ($179) to buy (they did not buy it); one more was excluded because the supplier went out of business and there were no services; and the last one was excluded because the content of the application was the same as that marketed by another company. Therefore, six were left.

Participants of experiment: Thirty adults aged 21 or above whom we recruited via classified advertisement, etc. 47% male, 53% in their twenties, 37% novices (self-assessment). They could use a Windows personal computer, had less than three hours of experience in using Access, and were not working as a programmer/Web developer (checked in the preliminary examination). They paid them $40 for their volunteer work, which took two hours.

Experimental process: They let each of participants experience only two out of six applications for the sake of making comparison. As there are a total of 30 combinations; i.e. six applications multiplied by five applications remaining after the first application was chosen, each volunteer was assigned one of the combinations. Allocation was done in such way that the effect of order was cancelled out. They preformed two sets of experiments, each consisting of 45 minutes' experience of the application, followed by posttest and a questionnaire. For experience of the application, at the start of each session they instructed, "Please try to look around and see as much as possible within 45 minutes as if you are learning this software at home." The posttest consists of 20 knowledge-related multiple choice questions divided into two sets, 10 questions each, prepared by an assistant who checked all six applications. The effect of order was cancelled out.

Ratings (results)

(1) Which software would you buy (provided the price is the same)? The number of times selected: A:9, B:7, C:5, D:5, E:3, F:1 (A was rejected only once, when it was compared with B; F was chosen only once, when it was compared with E). A and B: Preferred applications, E and F: Not preferred, C and D: In the middle [Figure 1]

(2) How big is the difference between the two applications? (1: small – 7: large): The average of 30 combinations = 5.3

(3) How do you rate various aspects of the application? (Aspects: Overall, ease of use, readability, attractiveness, clarity of objectives, manner of teaching, entertainment characteristic, navigation) → Those applications for which the rating was low (E and F) have extremely low entertainment-characteristic (Table 1).

Ratings and prices: Inversely proportional (Prices: A $20, B $27, C $55, D $98, E $75, F $99) (Table 5) → The cheaper the software, the more they liked it.

Ratings and performance: Proportional (Average score: A 8.4, B 7.8, C 6.7, D 7.3, E 7.5, F 6.3) (Table 1) The maximum score is 10 → It is NOT that the volunteers liked those applications that are high in entertainment characteristic and low in learning characteristic. Actually, rather the opposite was true. Those applications that were liked better scored good marks. [Considering the experiment procedure, it is also possible to understand that the volunteers liked it because the performance was good].

Characteristics of the application (Where does the difference come from?): Analysis of the results of the questionnaire and the characteristics

- They liked those applications that use a lot of graphic images and that are, in terms of structure, simple and easy to understand.

Software A is of a type in which animated characters talk to the user in unidirectional mode. Most of the options are not used. Software B, on the other hand, uses a lot of interaction. The volunteers liked the quiz game and the exercise corner → They have different preferences when it comes to interactivity. Being unidirectional is not necessarily bad (particularly for a beginner).

- On the other hand, those applications which offer a lot of text-based information and have a wide variety of functions were NOT very well liked.

Options of software F (10 minutes' of explanation of how to use is included): Exercises in a number of formats, quick check, reference, learners community, tutor, supplementary reference in PDF format, bookstore, software shopping page, help page, thread type BBS, and chat corner. Although the largest number of functions is offered and most sophisticated ID techniques are used, sometimes they are not used but are ignored. → Further research is required to find out whether this is only true for participants of experiments who tried it out in a mere 45 minutes and were paid, or if this also holds true for learners who use it for longer periods and are more motivated.

Difference from expert ratings: Except for D, for which the results of the two evaluations were the same, the ratings of other five applications were totally opposite. (Figure 5) → This might be an example of the bafflegab phenomenon. The study argues that one thing is for sure: that formative evaluation, which is stressed in the ID process, is important.

I have heard such stories as expert evaluation and the result of the trial experiment are opposite. Maybe this is also true in the world of e-Learning. The experiment is well constructed and the way they conducted the study is also suggestive.

The study introduced above

Hassett, J., Ingram, A., Hassett, M., & Marino, E. (2003). What Do Learners Like? Ratings of Off-the-Shelf Web-Based Training Courses. *International Journal on E-Learning, 2* (1), 50-60. [Available Online]: http://www.aace.org/DL/index.cfm/fuseaction/View/paperID/11553

References

Kajita, E. (1992) "Educational evaluation (2nd edition)", Yuhikaku [In Japanese]

The Mechanical Social Systems Foundation (2001) "Study on promotion of distance learning system" (CAIT) [Available Online] http://www.jipdec.jp/chosa/caittext/wbt/indexh12.htm [In Japanese]

Suzuki, K. (1989), 'Establishment of drill-based CAI which supplements the foreign language education programs on TV', "Educational broadcasting study" No. 17, pp. 21-37 [Available Online] http://www.iwate-pu.ac.jp/home/ksuzuki/ resume/journals/1989b.html

Suzuki, K. (2002), "Instructional Material design manual", Kitaohji-shobo [In Japanese]

Suzuki, K. (1989), "12. Use examples to explain how to clarify the instructional objectives", Kazuo Numano and Shigeru Hirasawa, "Educational methods/technologies", Gakubunsha [In Japanese]

Suzuki, K. (1987), "Regarding formative evaluation technique in the design and development of CAI material", "Audio-visual education study" No. 17, pp. 1-15 [In Japanese]

Dick, W., & Carey, L. (2001). *The systematic design of instruction* (5th Ed.). Scott, Foresman, and Company.

Flagg, B. N. (Ed.). (1989). *Formative Evaluation for Educational Technologies*. Lawrence Erlbaum Associates.

	End of chapter report assignment (Chapter 3)	

Please write a report on one or more of the following three assignments:

1) Summarize your questions, comments, opinions, and impressions you had after reading through this chapter (Chapter 3). In addition, if you have any experience or additional information or have done any research (do not forget to name the source) in relation to what is written in this chapter, you are encouraged to include them in your report so that you can extend your understanding even further.

2) Analyze an example of e-Learning which you know in relation to evaluation of corporate education in Japan while mainly consulting Section 2 of this Chapter. You are required to offer your opinion regarding what is true and to what extent that is true in relation to the background of the fact that they rarely carry out evaluation of corporate education in Japan, the reason why evaluation is necessary from now on, and the concept that the learners are customers. In addition, not limited to e-Learning examples, you are also encouraged to analyze your experience in terms of how you have been educated in school and at work, or in educational activities you are doing now.