

## CHAPTER II

### LITERATURE REVIEW

In this chapter, since this study aims at evaluating a task-based course, it is necessary to review and discuss the key concepts in relation to task-based approach, self-directed learning and learner autonomy, as well as, curriculum (or program/course) evaluation. The discussion will include previous research in language program evaluation.

#### **2.1 Task-based approach**

##### **1. Task-based instruction**

Due to the emergence of English as a global language on policies and practices in the Asia-Pacific region in the last fifteen years, the task-based approach has been claimed as a central principle underpinning English language curricula in this region (Lynch, 1996). 'Task-based learning', 'task-based instruction', 'task-based language teaching', and 'task-based language teaching and learning' are interchangeable terms used to refer to an alternative approach to language teaching which requires learners to learn a particular language by completing a given task. It is assumed that transacting tasks in this way will engage naturalistic acquisitional mechanisms causing the underlying interlanguage to develop (Skehan, 1998). A range of approaches to task-based instruction are evident in many research studies (Candlin, 1987; Nunan, 1989; Skehan, 1992, and Pica, Kanagy, and Falodun, 1993 cited in Skehan, 1996). They can be categorized into so-called 'strong and weak forms' of the task-based approach based on task characteristics. A strong form argues that tasks should be the unit of language teaching to which everything else is subsidiary. Tasks should be transacted to drive forward language development, resulting in second language acquisition that is similar to the process of interaction in first language acquisition (Wells, 1985 cited in Skehan, 1996). A weak version claims that tasks are an essential part embedded in a more complex pedagogic context. They may be preceded and followed by focused instruction that is contingent on task performance. With regard to this version, task-based instruction is very closed to general communicative teaching, and also compatible with the traditional 3P's (presentation, practice, and production) approach, only with production based on tasks.

The term 'task' has been defined differently by several teachers and writers. For example, Williams and Burden (1997) view a task as 'any activity that learners engage in to further the process of learning a language', whereas, Estaire and Zanon (1994) suggest a broader definition. They divide it into two subcategories: communication tasks that focus on communication and enabling tasks that focus on form. Many other teachers and writers make a clear distinction between tasks and exercises. Tasks involve communicative language use in which the learner's attention is focused on meaning rather than linguistic structures. Activities focusing upon practising specific elements of knowledge, skills and strategies needed for the task are called exercises. Skehan's (1998) definition of task supports this distinction:

- meaning is primary;
- there is some communication problem to solve;
- there is some sort of relationship to real-world activities;
- task completion has some priority;
- the assessment of task is in terms of outcome.

In addition, Willis (1998 cited in Skehan, 1998) describes that tasks:

- do not give learners other people's meaning to regulate;
- are not concerned with language display;
- are not conformity-oriented;
- are not practice-oriented;
- do not embed language into materials so that the specific structures can be focused on.

Besides, Nunan (2004) classifies tasks into real-world or target tasks, and pedagogical tasks. The former refers to uses of language in the world beyond the classroom. Examples of real-world tasks are painting a fence, dressing a child, filling out a form, buying a pair of shoes, etc. These are the things people do in everyday life. The latter, on the other hand, refers to those that occur in the classroom. For example, drawing a map while listening to a tape, listening to an instruction, and performing a command. Such tasks may or may not involve the production of language. Most definitions of pedagogical tasks emphasize that they involve communicative language use in which the user's attention is focused on meaning rather than grammatical form. Nunan (*ibid.*) portrays his different view on this issue.

He believes that meaning and form are highly interrelated. Grammar can facilitate language users to express different communicative meanings. Willis and Willis (cited in Nunan, 2004) point out that tasks are different from exercises in that learners are free to use a range of language structures that are not prespecified to achieve the task outcomes.

In evaluating a task, three different perspectives should be taken into consideration: tasks as workplans, tasks in process and tasks as outcomes. In other words, a task evaluation can be done by investigating the various types of information: (1) information about how the task was performed, (2) information about what learning took place as a result of performing the task and (3) information about the teachers' and the students' opinions about the task (Ellis, 1998).

## 2. Implementing task-based instruction

Two contrasting approaches to implementation of task-based instruction which differ in the way they relate tasks to specific language structure, are a structure-oriented approach and a communicative-driven approach. The first approach focuses on the inclusion of language structures in task design. Loschky and Bley-Vroman (1993 cited in Skehan, 1998) propose their approach that distinguishes between three structure-to-task relationships:

1. Naturalness: the use of a structure during a task would be unforced, i.e. would not stand out, but where alternative structures would do equally well.
2. Utility: the use of a particular structure would help the efficiency of the completion of the task, but could be avoided through the use of alternative structures or perhaps through the use of communication strategies.
3. Essentialness: a particular structure has to be used in order to complete a task.

In addition, Fotos and Ellis (1991 cited in Skehan, 1998) report their study in which specific structures were forced by particular tasks. Both cases illustrate that implicit learning materials can be used with some emphasis on the artificial transformation of materials so that particular structures become salient. However, Willis (1993,ibid.) argues that tasks cannot be used to trap structures without becoming unnatural.

The second approach is advocated by Willis (ibid.) who proposes that learners can develop language effectively through the use of natural tasks, which are not conformity-based or display-oriented for any particular structure. He points out that once tasks are transacted in an unforced way, they will generate the most significant lexis of a language. The learners will be able to use such lexis in syntactic patterns. Long (1993 cited in Skehan, 1998) is another supporter of this approach. He proposes that a task-based approach, in which real-world needs force task design, will generate interactions that activate acquisitional processes and develop interlanguage. However, this approach cannot deal with two crucial problems: (1) overemphasizing communication results in increased use of communication strategies and lexically-based language, and (2) there is no easy means of assuring systematic language development.

However, an intermediate approach (Long, 1988 cited in Skehan, 1998), which is claimed as a very middle-of-the-road one, proposes that:

- engaging in worthwhile, meaningful communication may emphasize fluency at the expense of the focus on form that accuracy and complexity would require;
- attentional limitations constrain the capacity of the language learner to focus on a number of different areas simultaneously;
- different aspects of language performance, particularly accuracy, fluency, and complexity, enter into competition with one another;
- using (properly defined) tasks to generate communication means that the use of particular structures can at best be probabilistic, and is certainly not guaranteed.

Besides, another two models for balancing concern for communication and form are proposed by Willis (1996) and Skehan (1998). Willis (1996) describes five principles for the implementation of a task-based approach:

1. There should be exposure to worthwhile and authentic language.
2. There should be use of language.
3. Tasks should motivate learners to engage in language use.
4. There should be a focus on language at some points in a task cycle.
5. The focus on language should be more or less prominent at different times.

Willis (1996)'s task-based learning framework consists of three phases: pre-task, task cycle and language focus (see Figure2.1).

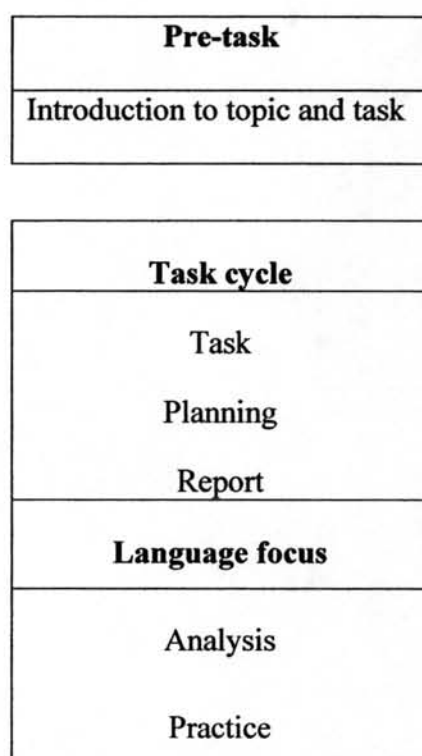


Figure 2.1: Task-based learning framework (Willis, 1996: 52)

Interestingly, Willis (1996) suggests that there may be some degree of explicit focus on a particular aspect of the language system, but it should come after a task has been done (and even after the reporting has been done). Any language focused upon is relevant to learners and required for a communicative purpose, rather than being introduced because a syllabus dictates that it should be covered at a particular point. Skehan (1998: 128-129) states that Willis's approach does have some drawbacks. First, there is no clear connection with theories on second language acquisition. Second, there is no explicit connection with research. Finally, this approach does not provide sufficient detail on the nature of second language performance, and the ways in which different pedagogic goals (fluency, accuracy, and complexity) can be achieved.

Skehan (*ibid.*, p. 129) proposes his five principles for task-based instruction as well as a model that he claims to be more theory-grounded, more systematic and more effective in promoting language development. The five principles are:

1. Choose a range of target structures.
2. Choose tasks which meet the utility criterion.
3. Select and sequence tasks to achieve balanced goal development.



4. Maximize the chances of focus on form through attentional manipulation.
5. Use the cycle of accountability.

Later, Skehan (*ibid.*, p. 139-140) proposes his model consisting of three main parts: the pre-task phase, the during-task phase, and the post-task phase. In the pre-task phase, three major types of activities include teaching, consciousness raising and planning. Teaching is concerned with the introduction of new language, and with restructuring. It may be deductive and explicit or inductive and implicit. The deductive approaches introduce new language to the interlanguage system, whereas the inductive approaches achieve restructuring of the underlying system. Consciousness raising activities, such as text exploration activities, pre-task brainstorming activities, and pre-task discussions can change the learner's awareness of elements of the task before it is done. As a result, the task is approached and performed differently. Consciousness-raising activities can also be used to reduce cognitive complexity. In pre-task activation sessions, learners can recall schematic knowledge relevant to the task. This will change their cognitive familiarity with the task. Planning time can produce the balance of pedagogical goals: fluency, accuracy, and complexity. However, given suggestions for both language and content, learners seemed to focus more on the content that causes more language complexity.

In the during-task phase, two general aspects are the manipulation of attention and pedagogic decisions. Skehan (*ibid.*) states that time pressure, modality, support, surprise, control, and stakes affect the communication stress. The speed with which a task needs to be completed is affected by time pressure. It is assumed that the greater the time pressures, the less attention can be paid to form either in terms of accuracy or complexity. Modality is concerned with a task form, whether it is spoken or written. The learners seem to pay less attention to form when doing spoken tasks due to the pressure of real-time performance. To facilitate students in doing a task, the teacher can use any kind of support, especially in the during-task phase. Teacher support is believed to make a task more structured and contain fewer errors. Surprise has no effect on learner performance in all aspects: fluency, accuracy and complexity. It is also possible for learners to take control over the way the task is done. They may be able to negotiate meaning, or even the nature of the task itself with the teacher. Some learners may prefer to avoid making errors and pay too much attention to a particular

structure. The teacher can help reduce the learner's attachment to form by explaining that task completion itself is linked to wider pedagogic goals.

With regard to pedagogic decisions, two proposals made by Willis (1996) and Sumuda et al. (1996 cited in Skehan, 1998) share some common features. Willis suggests that the during-task phase consists of three parts: (1) doing the task, (2) planning, and (3) reporting. Samuda et al. propose what they term 'knowledge-constructing tasks' in which meaning-based and form-based activities are combined. Such tasks do not provide any specific structures. A focus of language will be needed for the task to be completed successfully. Doing this, the teacher can supply form only when learners feel they need it. Samuda et al. emphasize that task design plays a critical role in structure-based tasks as they create scaffolding that highlights relationships between form and meaning in various ways and at various points within the task. Even though each proposal uses tasks which are purposeful, meaning-oriented, and so on, both of them value the importance of form-meaning links.

In the post-task phase, two important issues that influence the effectiveness of task-based instruction include: (1) altering attentional balance, and (2) encouraging consolidation and reflection. The first issue is concerned with the use of post-task activities to allocate learners' attention to form. The second issue aims at encouraging learners to restructure, and use the task and its performance as input to help in the process of noticing the gap between form and meaning, and developing language to cope with the shortcomings in the underlying interlanguage system.

In sum, task-based approaches have been developed to provide learners opportunities to do the tasks actively and to give an environment which promotes the natural language learning process. Examples of tasks are problem-solving, discussions, narratives and especially information-gap exercise in which learners have to transfer information to a partner who does not have it (Pica et al. 1994 cited in Skehan, 1998). Despite its tremendous benefits, a task-based approach to teaching yields some danger. The focus is on meaning rather than form. Students may be led to use fluent but inaccurate language. It is easy to see students use ill-formed words or phrases and some gesture and intonation when doing a task. Also, it is quite challenging for a teacher to choose, sequence, and implement tasks to promote both 'meaning' and 'form'. i.e., fluency and accuracy in language learning. Therefore, in

evaluating the effectiveness of a task-based course, both 'meaning' and 'form' should be examined.

### 3. Task-based syllabus

Task-based instruction has a close relationship with communicative language teaching. CLT is a broad, philosophical approach to the language curriculum that draws on theory and research in linguistics, anthropology, psychology and sociology, while task-based language teaching represents a realization of this philosophy at the levels of syllabus design and methodology (Nunan, 2004).

Wilkins (cited in Nunan, 2004) pinpoints the distinction between two approaches to syllabus design: 'syntactic' and analytical approaches. Syntactic approaches are regarded as traditional ways of organizing the syllabus. In syntactic approaches, discrete points of language are taught separately and step-by-step. The learner acquires in a gradual process until the whole structure of language has been built up. However, it seems that learners cannot acquire one item perfectly at a time. They learn numerous items imperfectly and simultaneously, as a result, their learning is unstable. Wilkins (ibid.) offers an alternative to syntactic approaches known as analytical approaches in which the learner is presented with holistic chunks of language and is required to analyze them, or break them down into their constituent parts. Task-based syllabuses belong to this category, as well as project-based, content-based, thematic-based, and text-based syllabuses. All of them share one thing in common—they do not rely on prior analysis of the language into discrete points.

A task-based syllabus is, therefore, not organized around linguistic features of the language being learned but according to some other organizing principles. In task-based instruction the content of the teaching is a series of complex and purposeful tasks that the students want or need to perform with the language they are learning. The tasks are defined as activities with a purpose rather than language learning. Language learning is subordinated to task performance, and language teaching occurs only as the need arises during the performance of a given task. Tasks integrate language (and other) skills in specific settings of language use. The goal of task-based teaching is to draw on resources for students to complete some piece of work- a process. Students draw on a variety of language forms, functions, and skills often in an individual and unpredictable way when completing the tasks. Tasks that can be used for language learning are generally tasks that the learners actually have to perform in any case. Examples would include applying for a job, talking with a social



worker, getting housing information over the telephone, completing bureaucratic forms, collecting information about preschools to decide to which one to send a child, preparing a paper for another course, and reading a textbook for another course.

It is worth noting that task-based language teaching and learner autonomy can be related in that they both regard learners as a centre in learning. Wenden (1991 cited in Tudor, 1996: 18) describes the autonomous learner as one who has acquired the strategies and knowledge to take some responsibility for her own language learning and is willing and self-confident enough to do so. Learning by doing is one of the key concepts of task-based learning that claims that learners learn best through doing- through actively constructing their own knowledge rather than having it transmitted by the teacher (Nunan, 2004: 34). Self-directed learning is a learning mode that encourages students to work independently with little teacher support. In task-based language teaching, learners should be given opportunities to reflect on what they have learnt and how well they are doing. Training on learning strategies, particularly metacognitive strategies, should be provided to learners to enable them become reflective learners.

## **2.2 Self-directed learning and learner autonomy**

Autonomous learning is one of the key concepts of the evaluated course. The terms 'self-directed learning' and 'learner autonomy' are closely related. However, there are few references of the term 'self-directed learning' in language learning. It tends to be used differently by different teachers. It is claimed that there is to some extent an overlap between the terms 'self-directed learning' and 'autonomy'. Boud and Sidery (cited in Dickinson, 1978) describe the two terms as "two names for the same phenomenon". They state:

We understand self-directed to imply maximum autonomy for the individual concomitant with concern for the autonomy of others, and the use of each other's resources in sensitive and effective ways.

Underlying this definition of self-directed learning is an assumption of what it means to be an educated person. The assumption is that an educated person is one who can identify his own needs, set his own goals, develop strategies for meeting his needs and be able to monitor his own action in this process. He

can co-operate with others to obtain mutual support and assistance so that each may gain fulfillment.

However, Dickinson (1978) argues that autonomy represents the upper limit of self-directed learning measured on a national scale from total direction to full freedom. Consequently, autonomy is one set of possibilities within the larger category of self-directed learning. In other words, where the learner's choices can be made freely, we have full autonomy. Where only some are freely made, we have some degree of autonomy.

In short, self-directed learning can be described as an approach which stresses the importance of individual differences, learner training and learner self-assessment. Learners are encouraged to choose what they want to learn or do that suits their learning styles. They can set their own goals of learning and choose means to achieve those goals. Moreover, they can monitor and assess their own work.

Autonomy is the ability to take charge of one's own learning. It is now believed that in order to develop learners' responsibility for their own learning, they need to obtain some training on learning strategies, and should know how to choose their own materials and how to evaluate themselves (Dickinson, 1995). Encouraging learners to develop autonomy is not just an ideal concept. Developing some degree of autonomy is essential if learners are to become effective language learners. Nunan (1997) argues strongly that autonomous learners are rare and fostering learner autonomy is best done in the language classroom. He outlines a scheme for gradually increasing the degree of autonomy exercised by learners in a programme of learning. The scheme consists of five levels for encouraging learner autonomy ranging from an initial level of raising learner awareness to the highest level of learners becoming fully autonomous. Littlewood (1996: 431) identifies three broad domains of autonomy:

1. autonomy as a communicator depends on (a) the ability to use the language creatively; and (b) the ability to use appropriate strategies for communicating meaning in specific situations;
2. autonomy as a learner depends on (a) the ability to engage in independent work (e.g. self-directed learning); and (b) the ability to use appropriate learning strategies, both inside and outside the classroom;

3. autonomy as a person depends on (in the foreign language learning context) on (a) the ability to express personal meanings; and (b) the ability to create personal learning contexts, e.g. through interacting inside the classroom.

In the applied linguistics literature *autonomy* is viewed as a capacity for active, independent learning, critical reflection and decision making (Dickinson, 1995: 167). Littlewood (1996: 428) explains that the independent capacity of a person to make and carry his or her own choices depends on two main components: *ability* and *willingness*. Ability depends on possessing both *knowledge* about alternatives from which choices have to be made and the necessary *skills* for carrying out whatever choices seem most appropriate. Willingness depends on having both *motivation* and the *confidence* to take responsibility for the choices required. To become autonomous, these four components need to be present together. In practice, these components are closely linked. "The more knowledge and skills the students possess, the more confident they are likely to feel when asked to perform independently; the more confident they feel, the more they are likely to mobilise their knowledge and skills in order to perform effectively" (ibid.).

Moreover, Wang and Palincsar (1989 cited in Dickinson, 1995: 168) claim that motivation to learn and learning effectiveness can be increased in learners who take responsibility for their own learning, who understand and accept that their learning success is a result of effort, and that failure can be overtaken with great effort and better use of strategies. Learners who are interested in the learning tasks and the learning outcomes for their own sake tend to have higher motivation than those interested in rewards resulting from success. In other words, motivation tends to be higher in learners who focus on learning outcomes rather than performance outcomes (Deci and Ryan, 1985; Dweck, 1986 cited in Dickinson, 1995: 168).

Later, Dickinson (1995: 171) proposes his attribution theory stating that the learner's perception of the cause of his or her own success or failure has influence on perceptions of future performance. Four possible causes categorized according to stability (i.e. whether it can be changed or not), internal or external to learner, and whether the learner can control it include:

1. Ability (internal and stable)
2. Task difficulty (external and stable)

3. Effort (internal, changeable and under the learner's control)
4. Luck (external, changeable but not under the learner's control)

Pupils who attribute their failure to stable causes— ability and task difficulty—tend not to persist when they fail, but those who believe that their failure is due to unstable or internal causes— particularly effort, tend to persist in the face of failure (ibid.) It can be said that children's views on the stability of intelligence have strong effects on their learning. Dweck (1986 cited in Dickinson, 1995) reports that children's theories of intelligence orientate them towards different goals. Children who believe that intelligence level is fixed tend to be concerned with performance goals, while children who believe that intelligence can be changed and increased tend to be concerned with learning and learning goals. Success enhances motivation only in children who are focused on learning goals (i.e. who are intrinsically motivated). Dickinson (ibid.) claims that attribution theory relates to learning autonomy in that it provides evidence to show that learners who believe that they have control over their learning—that by accepting new challenges they can increase their ability to perform learning tasks and so increase their intelligence—tend to be more successful than others.

It is interesting to note that some form of self-directed learning with institutional support in the shape of counselling and resource centres has been found very useful in numerous institutions. Self-directed schemes have been regarded suitable not only for English as a Foreign Language, but also for other foreign languages as well. A large number of courses in French and German were provided by many Eurocentres. The Language Centre of University of Cambridge offers self-directed learning for more than 40 different languages (Gremmo and Riley, 1995). However, there is no universal model for setting up a self-directed scheme. It can be adapted to various institutional requirements and expectations, to the particular characteristics of the learners and the staff, and to meet different local needs.

In order to search for the evaluation model that can evaluate a task-based English course, the theoretical framework of program evaluation should be taken in account.

### **2.3 Curriculum (or program) evaluation**

#### **1. Definitions**

Evaluation has existed for more than two centuries. The very first instance



was of civil service examinations conducted by Chinese officials to measure the proficiency of applicants for government positions (Fitzpatrick, Sanders and Worthen, 2004). Evaluation is still vital in any social and educational program. In the language teaching field, especially in ESL/EFL programs, the term *evaluation* is generally defined in terms of its purposes and the types of information to be gathered and used. The most important purpose of evaluation is to make decisions on the worth of a particular program. To be more specific, evaluation is conducted to find any relevant information needed for the improvement of a curriculum and to assess its effectiveness within the particular contexts. Richards et al. (1985 cited in Brown, 1995) defines evaluation as “the systematic gathering of information for purposes of making decisions.” Brown (Brown, 1995 cited in Brown, 1996) defines his definition of evaluation as “the systematic collection and analysis of all relevant information necessary to promote the improvement of a curriculum and analyze its effectiveness within the context of the particular institution(s).” Nunan (1992) makes a clear distinction between the terms *assessment* and *evaluation*. *Assessment* refers to the processes or procedures whereby we determine what learners are able to do in the target language (i.e., assessing learners’ language abilities). *Evaluation*, on the other hand, refers to a wider range of processes which may or may not include assessment data (i.e., evaluation of a course/program). Nunan describes evaluation as a decision-making process which involves two important characteristics: making value judgments and action. The data resulting from evaluation can pinpoint whether a course needs to be modified or altered in any way in order to achieve goals and objectives set for a course. Evaluation is an effective means of measuring teaching and learning performances in a language program and of improving the teaching process. It can be used to trace both teaching procedures and learning process (Yunian & Ness, 1999).

Besides, the term *evaluation* can be viewed differently according to the types of information collected and analyzed. Gasper’s definition of evaluation (1995 cited in Yunian & Ness, 1999) can illustrate this point, that is, “evaluation is a process to judge or measure the value of a finished or ongoing program, plan, or even a policy”. This definition classifies evaluation into two main types: process evaluation and product evaluation. In other words, the former is widely known as formative evaluation and the latter as summative evaluation. However, these two bipolar concepts tend to interact each other. Graves (2000) points out that course evaluation

answers the question *How effective is/was the course in helping learn a target language?* An evaluation plan includes feedback on the course and a summative course and teacher evaluation.

In short, from the various definitions of evaluation mentioned above, evaluation can be viewed as a systematic means of gathering, analyzing, and synthesizing information— both of “process” or “product”, for the purposes of making value judgements: retaining or improving the existing curriculum and promoting teaching and learning process required by the curriculum.

There have been strong arguments about evaluation as research. Some researchers believe that evaluation cannot be considered a form of research as most evaluations making judgements about a single program lack external validity. Nunan (1992:193) argues that any investigation which contains questions, data, and interpretations of the data qualifies as research. He also elaborates his advocacy of evaluation studies.

I would therefore accept that evaluations, even those of a single program, are, in fact research. While evaluators who are investigating a single program can usually ignore issues relating to external validity, they still have a responsibility to guard against threats to the internal and external reliability, and so also the internal validity of their investigations. Internal validity can be particularly problematic for research conducted in a field setting.

Due to the concern about external validity (i.e. generalizability) of evaluation studies, many evaluators in the field of applied linguistics in particular tend to favor a traditional, quantitative experimental approach to conducting research known as the positivistic paradigm (e.g., Campbell and Erlebacher, 1970; Cain, 1975; Campbell and Boruch, 1975; Cronbach et al., 1975; Kenney, 1975; Bryk and Weisberg, 1976; Boruch and Rindskopf, 1977; Cook et al., 1977; Bryk et al, 1980 cited in Lynch, 1996). Another alternative approach to inquiry is the naturalistic paradigm that is very much opposed to the positivistic one in the sense that it emphasizes qualitative methods of inquiry including observing, describing, interpreting, and understanding how events take place in real-world settings. This approach views the educational program being evaluated as a process rather than product.

## 2. Approaches to evaluation

Different approaches have been developed for the task of program evaluation. Brown (1995) suggests four approaches to program evaluation: (a) product-oriented approaches; (b) static-characteristic approaches; (c) process-oriented approaches; and (d) decision-facilitation approaches.

### 2.1 Product-oriented approaches

Product-oriented approaches focus on the evaluation of the goals and instructional objectives with the purpose of determining whether they have been achieved. The primary advocates of this approach were Tyler, Hammond, and Metfessel and Michel.

Tyler (1942 cited in Brown, 1995) put much emphasis on the evaluation of the so-called 'measurable behavioral objectives'. He pointed out that if such objectives of a particular program have not been achieved, that program has failed to attain its goals. In addition, the development of goals and objectives should be based on a number of sources: the students, the subject matter, instructional materials, the society at large, philosophy of education, learning philosophy, and so on.

In addition, Hammond's model of curriculum evaluation (cited in Brown, 1995), which was also product-oriented, consisted of five steps:

1. Identifying precisely what is to be evaluated
2. Defining the descriptive variables
3. Stating objectives in behavioral terms
4. Assessing the behavior described in the objectives
5. Analyzing the results and determining the effectiveness of the program.

Besides, Metfessel and Michael (1967 cited in Brown, 1995) provided more detailed information which includes eight major evaluation phases:

1. Direct and indirect involvement of the total school community
2. Formation of a cohesive model of broad goals and specific objectives
3. Translation of specific objectives into communicable form
4. Instrumentation necessary for furnishing measures that allow inferences about program effectiveness
5. Periodic observations of behaviors
6. Analysis of data given by status and change measures
7. Interpretation of the data relative to specific objectives and broad goals

8. Recommendations culminating in further implementation, modifications, and revisions of broad goals and specific objectives.

Therefore, it is worth noting that these product-oriented approaches examine whether student behaviors meet the goal and instructional objectives of the curriculum. The success of the program can be measured in terms of the degree to which those objectives are achieved.

### 2.2 Static-characteristic approaches

The static-characteristic evaluation is also conducted to determine the effectiveness of a particular program. Evaluators are outside experts who inspect the program by examining various accounting and academic records, and such static characteristics including the number of library books, the number and types of degrees held by the faculty, the student-to-teacher ratio, the number and seating capacity of classrooms, the parking facilities, and so forth. Static-characteristic evaluations are currently used as a means for institutional accreditation. These approaches have been unsatisfactory to many educators due to its reliance on intrinsic rather than extrinsic factors as well as lack of empirical evidence to support that such factors are associated with the final outcomes.

### 2.3 Process- oriented approaches

These approaches are advocated by Scriven and Stake. Scriven's goal-free evaluation model is evaluation in which limits are not set on studying the expected effects of the program vis-à-vis the goals. Scriven distinguishes between formative and summative evaluation, and stresses that the evaluators must pay much attention not only to the final outcomes but also to the unexpected outcomes arising during the evaluation process. Stake's 'countenance model' is also process-oriented. Stake (1967 cited in Brown, 1995) suggests that evaluators must keep the differences between the two types of activities, descriptive and judgmental, in mind.

### 2.4 Decision-facilitation approaches

The ultimate aim of this type of evaluation is to make decisions. In the decision-facilitation approaches, evaluators must avoid making any judgments, but instead, gather as much information as possible for the administrators and faculty in the program to make their own decisions. Examples of this approach are the CIPP, CSE, and Discrepancy models of evaluation.



CIPP stands for Context, Input, Process, and Product, the four types of evaluation identified by Stufflebeam (1971 cited in Nunan, 1992). It was very influential in the 1970s and 1980s, and it is now still in fashion among many program evaluators. The CSE model is named after the acronym for the Center for the Study of Evaluation at UCLA. Alkin (1969 cited in Brown, 1995) points out that evaluation should provide information for five different categories of decisions: system assessment, program planning, program implementation, program improvement, and program certification. The discrepancy model advocated by Provus (1971 cited in Brown, 1995) consists of five stages of evaluation: program description stage, program installation stage, treatment analysis stage, goal achievement analysis stage, and cost-benefit analysis.

Also worth noting is the fact that each approach has its own strengths and weaknesses. For example, product-oriented and static-characteristic approaches determine the effectiveness of the language program by looking at the degree to which the behavioral objectives are achieved by the students, but overlooking any factors that may have effects on the students' performance. These two approaches are also categorized as summative evaluation, which should be carried out at the end of the program. Due to the shortcomings of the product-oriented and static-characteristic approaches, a process-oriented approach is designed to pay more attention to the so called 'unexpected outcomes', i.e. other possibilities that might happen during the evaluation process. The process-oriented approach is formative and goal-free, and seems to be more dynamic than the first two approaches. However, the engagement of the evaluators in both descriptive and judgement activities in the process-oriented approach is opposed by some educators. A decision-facilitation approach is an alternative for those who believe that evaluators should be information gatherers, rather than decision-makers, in the process of program evaluation.

More recent approaches have been suggested by Owen and Rogers (1999). They connect their evaluation approaches: objectives-based, needs-based; and goal-free evaluation, with what they call evaluation forms classified into five categories: (1) proactive; (2) clarificative; (3) interactive; (4) monitoring; and (5) impact.

Proactive evaluation is a type of evaluation which takes place before a program is designed. The purpose is to help program planners make decisions about what type of the program is needed and how to develop an effective one. The evaluator is an advisor who provides evidence about policy development, program

format and procedures on how to change the organisation to make the program more effective. The approaches consistent with this form are needs assessment, research review and review for best practice and the creation of benchmarks.

Clarificative evaluation concentrates on clarifying the theory or logic of a program that has not been fully specified even though the program is being carried out. It refers to the casual mechanisms understood to link program activities with intended outcomes. In this form of evaluation, the evaluator or the program staff can conduct the program clarification. In collecting and analysing data, interview, observation and document analysis are involved. Approaches consistent with this form are logic development or evaluability assessment (i.e. the development of program logic using a range of analytical methods including documentation, and interviews with program staff and other stakeholders with a view to constructing a map of what the program is intended to do), and accreditation (i.e. a determination of the worth of program guidelines, generally in the context of certifying that an agency or organization can deliver the program for a given period).

Interactive evaluation is concerned with providing information about delivery or implementation of a program or about selected component elements or activities. It helps staff to more fully understand how and why a program operates in a given way. It involves middle-level managers and program implementers in the evaluation process. Approaches which are consistent with this form include responsive evaluation (i.e. the documentation or illumination of the delivery of a program), action research (i.e. the determination on how innovatory approaches to delivery are making a difference), quality review (i.e. institutional self-study), developmental evaluation (i.e. a continuous improvement process of a program), and empowerment evaluation (i.e. the development and evaluation done by program providers and participants of a program).

Monitoring evaluation is appropriate with a well-established ongoing program. It involves the development of a system of regular monitoring of the program's progress by using a range of both quantitative and qualitative techniques. It involves internally located evaluators who can utilize the following approaches: (1) component analysis that involves the systematic evaluation of a component of a large-scale program; (2) developed performance assessment that involves the development by which an organisation or system sets up evaluation procedures by which components can report regularly on their progress; (3) systems analysis that involves setting up

procedures by which the central management institutes common evaluation procedures to be used uniformly across an organisation.

Impact evaluation is used to assess the impact of a settled program. It can be described as a summative evaluation that assists in determining whether to terminate or whether to adopt it in another place. Approaches consistent with this form are:

1. objectives-based evaluation: judging the worth of a program on the extent to which the stated objectives of the program have been achieved
2. process-outcomes studies: measuring the degree of implementation of the program as well as determining the outcomes
3. needs-based assessment: judging the worth of a program on the extent to which the program has met the needs of the participants
4. goal-free evaluation: judging the unintended outcomes of the program

Obviously, Owen and Rogers (1999)'s forms-related evaluation approaches provide a wider range of roles for evaluation as well as more practical directions for the conduct of 'real' evaluations in the field.

#### **2.4 The responsive approach**

In 1967, several evaluation theorists expressed their concerns over the preoccupation of traditional evaluation approaches such as Tyler's objectives-oriented model which required evaluators to state and classify objectives, design elaborate evaluation systems, develop objective instrumentation, and prepare long technical reports. Consequently, they proposed a new orientation to evaluation known as participant-orientation approach. Robert Stake was the first evaluation theorist who advocated this approach. His paper "The Countenance of Educational Evaluation" focusing on the judgment of participants, brought about a dramatic change in the thinking of evaluators who did not believe that the products of the large-scale evaluations, such as numbers, figures, charts and tables, could really describe the phenomena being studied (Fitzpatrick, Sanders and Worthen, 2004).

During the early 1970s, Stake expanded his earlier work into a less formal but more pluralistic and process-focused approach which he named responsive evaluation. He defined his approach as follows:

An educational evaluation is responsive evaluation if it orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspective present are referred to in reporting the success and failure of the program (Stake, 1975).

Stake's responsive evaluation assumes value pluralism and facilitating judgment among local stakeholders. Stake emphasizes that there are different constituencies, different stakeholders who have different expectations, and different values. The evaluator should understand them and be in a good position to evaluate them, to represent them to readers and outsiders as part of the evaluation task (Abma, and Stake, 2001). He also adds that responsive evaluation is oriented around the important issues emerging in the evaluation context. To be responsive, the evaluator should come to know the circumstances, problems and values well, then use professional talent and discipline to carry out the inquiry. Stake believes that a major priority of a responsive evaluation's structure should be to provide the audience of the evaluation with a "vicarious experience" so that the audience has a better understanding of what the program is like. Stories and portrayals of people, places, and events are the most appropriate narrative form to create a vicarious experience, because their life-likeness and concreteness is close to direct, personal experience.

Stake also advocates case study methods in evaluation. He claims that knowledge from case studies is concrete and contextual, and open for different interpretations (ibid.). He calls his own specific type of case studies, *intrinsic* case studies of which the primary purpose is to understand a particular case rather than to investigate a certain issue (Stake and Mabry, 1995 cited in Abma, and Stake, 2001 ). The following is Stake's responsive evaluation plan divided into twelve recurring phases or events:



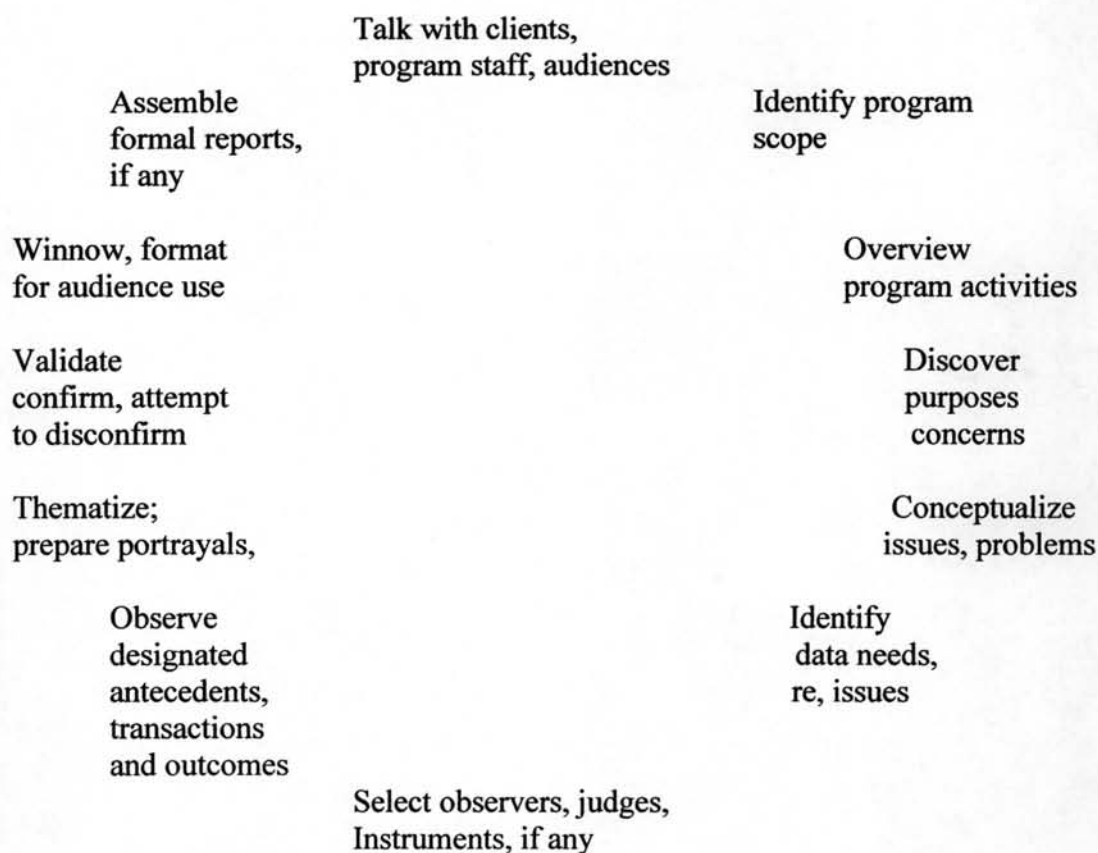


Figure 2.2: Prominent events in a responsive evaluation

Source: From *Program evaluation: Particularly responsive evaluation* (Occasional paper No. 5 By R.E. Stake, 1975, Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign.

Stake (1975: 19) likes to portray his twelve recurring events as if they are on the face of a clock. He claims that his clock can move clockwise, counter-clockwise, and cross-clockwise. In other words, any event can follow any other event. Many events may occur simultaneously. The evaluator may return to each event many times before the evaluation ends.

Later, Stake (1978 cited in Fitzpatrick, Sanders and Worthen, 2004) described his rationale for this approach as follows:

1. It helps audiences understand the program if evaluators pay attention to the natural way in which audiences understand and communicate about things.
2. Knowledge gained from experience (tactic knowledge) facilitates human understanding and extends human experience.

3. Naturalistic generalizations, arrived at by recognizing similarities between objects and issues in and out of context, are developed through experience. They serve to expand the way in which people come to view and understand programs.
4. By understanding single objects, people accumulate experiences that may be used to recognize similarities in other objects. Individuals add to existing experience and human understanding.

Among all the participant-oriented approaches, responsive approaches are regarded as the most powerful, flexible, and ready to use. They can be included in all other approaches. As a participant-oriented approach is not a 'pure' naturalistic inquiry, it can use both qualitative and quantitative methods. However, there is more use of qualitative methods in participant-oriented approaches than is typical in, for example, objectives or management-based evaluations. Using this approach, evaluators may see the potential for gaining new insights and usable new theories about educational, social, or corporate programs. Training users in evaluation as a part of the evaluation process, providing contextual variables, and encouraging multiple data collection techniques, contribute to the flexibility of this approach. It also provides rich, persuasive and credible information that reflects what really happens in a program. The greatest strength of this approach is that it empowers the quiet, powerless stakeholders to speak out on how they feel about the program as well as evaluation.

Subjectivity is the most serious limitation as this approach relies on human observation and individual perspective. Its tendency to minimize the importance of instrumentation and group data has been criticized as "loose and unsubstantiated". Intuitive data processing may cause bias in naturalistic evaluations. Ethnographic field work takes much time to complete. Moreover, this approach minimizes the central role evaluators should play in judging merit or worth. This approach can be labor-intensive, requiring full-time presence of the evaluator in the field over an extended period. As a result, only limited numbers of cases can be studied intensively resulting in the problem of generalization.

### **2.5 The Tylerian evaluation approach**

This approach, which was also known as objectives-oriented or objectives-

based evaluation, was developed by Ralph W. Tyler in the late 1930s. The Tyler rationale for evaluation evolved from two closely related rationales: one for achievement testing and the other for developing a curriculum and plan of instruction (Tyler, 1989). In his article entitled, *A Generalized Techniques for Constructing Achievement Tests*, Tyler described a procedure for developing an achievement test:

1. Identify the objectives of the educational program.
2. Define each objective in terms of behavior and content.
3. Identify situations where objectives are utilized.
4. Devise ways to present situations.
5. Devise ways to obtain a record.
6. Decide on the terms to use in appraisal.
7. Devise means to get a representative sample.

The rationale for curriculum development evolved from his experience as a director of evaluation for the eight-year study presented in the syllabus entitled, *Basic Principles of Curriculum and Instruction*. It identifies four basic questions that should be answered for developing a curriculum and a plan of instruction:

1. What education objectives are the students to be helped to attain? (i.e. what are they to be helped to learn?)
2. What learning experiences can be provided that will enable the students to attain the objectives? (i.e. how will the students be helped to learn?)
3. How will the learning experiences be organized to maximize their cumulative effect? (i.e. what sequences of learning and what plan of integration of learning experiences will be worked out to enable students to internalize what they are learning and apply it in appropriate situations that they encounter?)
4. How will the effectiveness of the program be evaluated (i.e. what procedure will be followed to provide a continuing check on the extent to which the desired learning is taking place)?

Obviously, this rationale reveals Tyler's concerns about three key features in evaluation: objectives, learning experiences and organization. The basic questions in the rationale are also viewed as parts of a cyclical procedure rather than a linear one. Tyler's rationale is considered logical, scientifically acceptable, and readily adoptable

by evaluators who advocate the pre-and-post-test measurement of behaviors (Fitzpatrick, Sanders and Worthen, 2004).

Tyler views evaluation as the process of determining the extent to which the objectives of the program are actually being attained (Fitzpatrick, Sanders and Worthen, 2004). The steps of Tyler's approach are:

1. Establish broad goals or objectives.
2. Classify the goals or objectives.
3. Define objectives in behavioral terms.
4. Find situations in which achievement of the objectives can be shown.
5. Develop or select measurement techniques.
6. Collect performance data.
7. Compare performance data with behaviorally stated objectives.

Even though Tyler emphasizes what he calls behavioral objectives, he advocates the use of general goals to establish purposes rather than a preoccupation with formulating them. Tyler (1991 cited in Fitzpatrick, Sanders and Worthen, 2004) emphasized the importance of the measurement of outcomes that it requires clear definitions of the desired patterns of behavior and of other possible outcomes both positive and negative. Evaluators need to choose or develop test situations that evoke such behavior from the students, use relevant criteria in appraising the students' reactions to these test situations, and report the results so that they can be understood by those who can use them constructively.

The objectives-based evaluation is very workable and defensible. It is a straightforward procedure of assessing learner achievement of objectives that determines success or failure and justifies improvements, maintenance, or termination of program activities (Fitzpatrick, Sanders and Worthen, 2004). Also, since it is easily understood, and easy to follow and implement, a number of evaluations have been done having goals and objectives as a central focus in the evaluation procedure. However, this approach does have some limitations:

1. It lacks real evaluative components (instead of making explicit judgment of merit or worth, it merely facilitates and assesses students achievement of objectives).
2. It lacks standards to judge the importance of observed discrepancies between objectives and performance level.



3. It neglects the value of the objectives themselves.
4. It ignores important alternatives that should be considered in planning a program.
5. It neglects the context in which the evaluation takes place.
6. It ignores the unintended outcomes of the activity other than those covered by the objectives.
7. It omits evidence of program value not reflected in its own objectives.
8. It promotes a linear, inflexible approach to evaluation.

In sum, the approach's overemphasis on test components tends to limit the evaluation's effectiveness and potential.

## **2.6 Integrated evaluation model**

In applied linguistics, there are two competing paradigms in program evaluation (the quantitative and qualitative debate). The positivistic paradigm refers to the traditional, experimental approach to evaluation. This approach has identified two major categories of research design: true experiments and quasi-experiments. The naturalistic paradigm is an alternative approach. The emphasis is on observing, describing, interpreting, and understanding how events take place in the real world rather than in a controlled, laboratory-like setting. The naturalistic evaluator gathers the evaluation data using techniques such as in-depth interviews, participant observation, and journals. The idea of compatibility between the two paradigms has been influenced by the notion of triangulation, a term introduced by Denzin (1970 cited in Lynch, 1996), for the collection of information from many different sources using different methods in order to avoid the bias inherent in any one particular source or method. Payne(1994 cited in Wilde and Sockey, 1995) describes four types of triangulation as:

- (1) using several different evaluators, with different orientations (e.g. qualitative and quantitative);
- (2) using several data sources (e.g. standardized tests, alternative assessments and interviews);
- (3) using several data collection methods (e.g. reviewing students' cumfolders and surveying teachers); and

- (4) using different theoretical approaches (e.g. using an evaluator familiar with and supportive of two-way bilingual education and another evaluator familiar with and supportive of transitional-type programs).

This research, therefore, utilizes multiple methods in the study of the same phenomenon to enhance the overall evaluation design. The objective-based approach representing the positivistic paradigm is integrated with the responsive approach of the naturalistic paradigm that seems to be the other side of the coin. The reason why the researcher selected these two approaches is the fact that the weakness of one particular design can be compensated by the strength of another design. Stake's (1975:20) comparison of responsive evaluation with preordinate evaluation (i.e. objective-based evaluation) in terms of the percentage of time evaluators of each approach spend on several evaluation tasks reveals what is needed or lacking in each.

	<i>Preordinate</i> (%)	<i>Responsive</i> (%)
Identifying issues, goals	10	10
Preparing instruments	30	15
Observing the program	5	30
Administrating tests, etc.	10	—
Gathering judgments	—	15
Learning client needs, etc.	—	5
Processing formal data	25	5
Preparing informal reports	—	10
Preparing formal reports	20	10

Therefore, it is thought that if the two approaches are combined in a single study, the credibility, reliability and validity of evaluation findings will be enhanced.

In his responsive evaluation, Stake widens the scope of evaluation dominated by Tyler's idea emphasizing a single comparison between intended and observed outcomes to responsive evaluation. He claims that Tyler's model emphasizes data-gathering of only two kinds of information: the goals and the outcomes. However, Stake admits that the evaluator does not need to change the research methods, e.g. quasi-experimental approaches or correlational surveys, but just incorporates additional data that are needed.

Based on Stake's responsive evaluation integrated with Tyler's objectives-based approach, the IST model is, therefore, different from other evaluation models in that it involves the stakeholders, students and program staff in particular, in the

evaluation process and makes multiple use of different measures. It is also different from the original models of Stake and Tyler (see Figures 2.3 and 2.4).

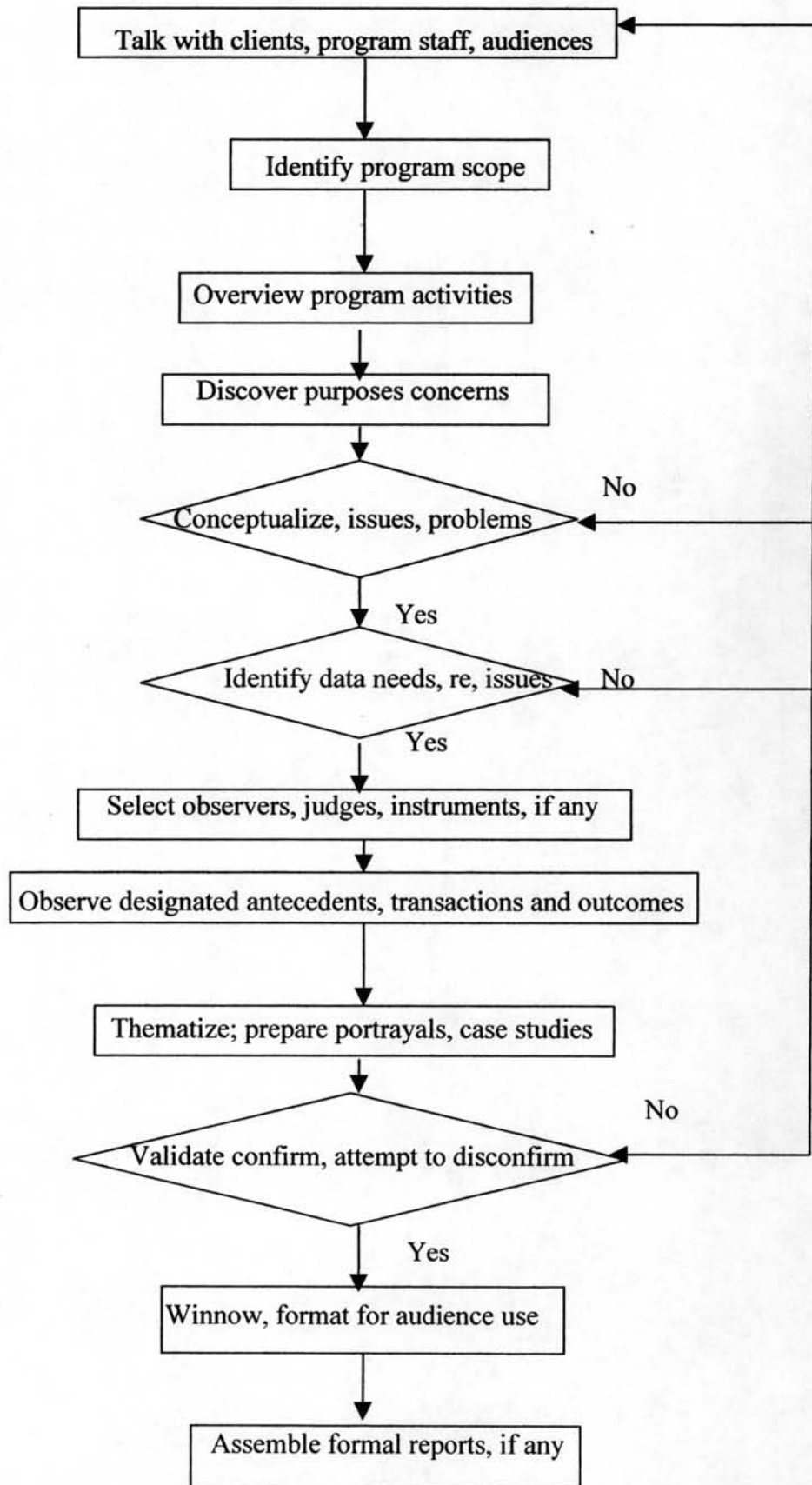


Figure 2.3: Stake's model



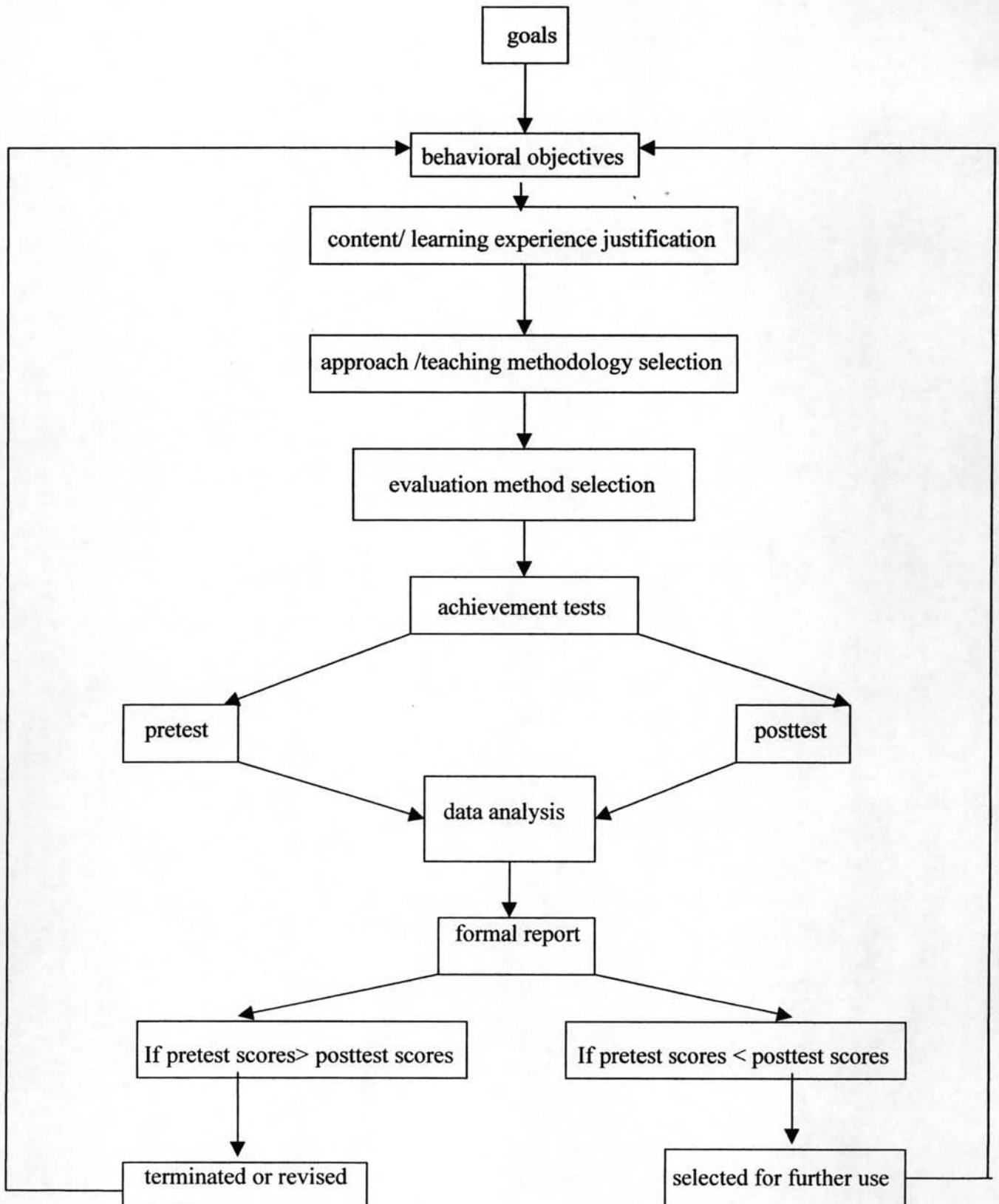


Figure 2.4: Tyler's model

The pretest/posttest model proposed by Tyler primarily focuses on student performance which can be measured objectively and explicitly. He uses discrepancies between what is expected and what is observed to reveal any program deficiencies. Interestingly, his approach can determine only 'what', but not 'how' the program objectives are attained.

On the contrary, the responsive evaluation evaluator seems to be more complex. The evaluator needs to do so many things simultaneously. He/she works cooperatively with stakeholders as well as audiences. After keeping records using different methods, the evaluator needs to ask for their feedback to confirm or disconfirm the findings. A final written report might be prepared depending on what the evaluator and the clients have agreed on. The evaluator should not rely on his own power of observation, judgement, and responding. Students, teachers, community leaders, curriculum specialists too should be involved in the evaluation.

The proposed model developed by the researcher of this study integrates the two models and includes some initiatives. Due to the purposes of this study that aims at examining students' language abilities and learning process, Tyler's pretest/posttest model is integrated with Stake's case study approach.

In Figure 2.5, the IST model is illustrated by using some colors as symbols. Blue represents Stake's responsive approach. Red represents Tyler's objective-based approach. Pink represents the researcher's initiatives.

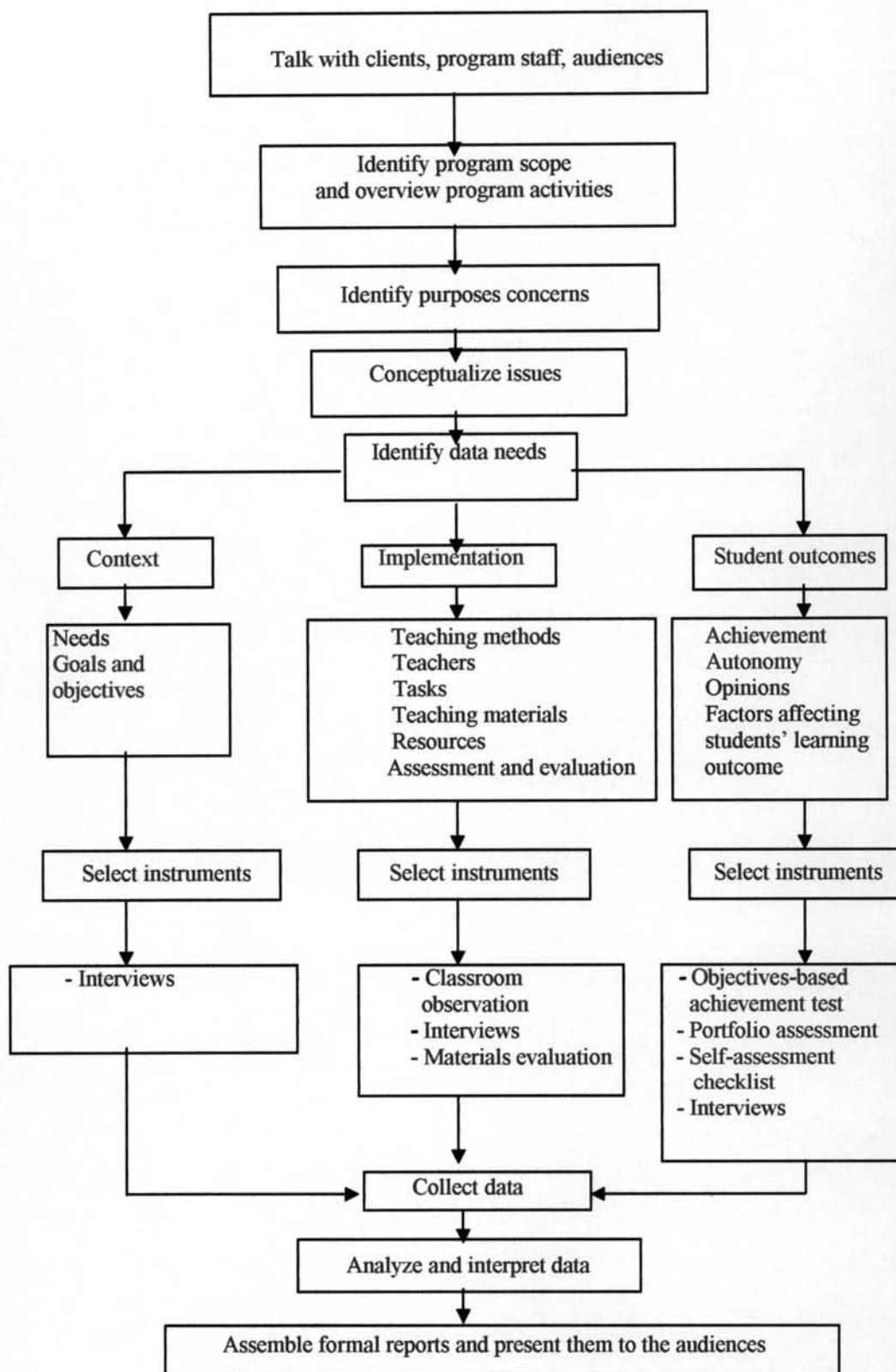


Figure 2.5: The Integrated Stake-Tyler model (IST)

The procedure starts with communication with the clients, the program staff and the audiences. The researcher needs to find out whether her ideas of the program scope, activities, purposes, and issues are consistent with those of the stakeholders. Then, the researcher identifies the program scope, purposes and issues to be studied, and then, the types of data needed, after which she selects the instruments. Qualitative and/or quantitative methods can be used depending on the types of data. After collecting the data, the researcher analyzes and interprets it both statistically and descriptively. The findings obtained from each method are assembled into formal reports and presented to the audiences.

In this model, both approaches are refined to fit a particular purpose as well as the setting of the study. Some parts of the two models are kept, and some are eliminated. Some new elements initiated by the research of this study are added.

As this research is a naturalistic inquiry, most of the plan follows Stake's responsive approach, i.e. responsive-evaluation procedures. This is because the ultimate goals of this study are:

1. to evaluate what is really happening in the program rather than what we would like to be happening.
2. to examine whether the evaluated program responds to the audience's requirements for information, and
3. to investigate different value perspectives of the people involved in reporting the success and failure of the program.

The English task-based course that will be evaluated has been used for some time, and modified several times; yet, the latest version cannot be claimed as the 'best'. The reasons for these changes are still in doubt. As a responsive evaluator, the researcher has to find out what the problems are, and whether there are some gaps in this curriculum. Talking with the stakeholders as well as the audience, such as students, teachers, program administrators, program directors, etc. facilitates the researcher to conceptualize issues or problems or potential problems that are a structure for continuing discussions with the clients, the staff, and the audience, and for a data gathering-plan. At the same time, a careful study of the evaluated course needs to be made, and its goals and objectives should be identified. At this stage, Tyler's model can be merged into Stake's, as they are similar in some points. Tyler



prefers to use general goals to establish purposes rather than a premature specification of behavioral objectives. In his model, the evaluator establishes broad goals and objectives, and then classifies them before defining the behavioral objectives. Stake's model requires the evaluator to get acquainted with a program as well as with people to get a holistic view of the program in order to search for 'issues'. After the issues have been identified, data needs and instruments would be identified and selected. Test and any data-gathering tools can be used depending on the program purposes. Responsive evaluation does not rule out any formal statements or abstract representations, e.g., flow charts, test scores, but they are treated not as the basis for the evaluation plan but as components of the instructional plan. As a matter of fact, instruments like observations, interviews, portrayals, etc., used in responsive evaluation are highly subjective. To overcome the worst aspect of subjectivity, more reliable and objective tools are needed. Tyler's achievement tests are selected as they are regarded the 'best' instrument to measure student outcomes.

In his model, Stake uses portrayals --descriptive accounts of a person, classroom, school, site, projects, activity, and so on—to provide *a vicarious experience* to the audience, as he believes it is a natural way in which people assimilate information and arrive at understanding. Some portrayals are short, such as a five-minute script, a log, or a scrapbook. A longer portrayal requires several media: narratives, maps, graphs, exhibits, taped conversations, photographs, and even audience role-playing. Some audiences should be selected to take part in the vicarious experience. The evaluator should present his portrayals, e.g., twenty displays of individual students who participate in case studies, to the audiences and ask them about the strengths and weaknesses of the program. The information gained will be included in his evaluation report. Portrayals are eliminated in this refined version because they are impractical, time-consuming, and highly subjective. Portfolios, a more recent tool which has more or less similar characteristics to portrayals, were selected as a substitute. Portfolios are a kind of prose description of each student and may include a variety of tasks; for example, student assignments, projects, reports on test scores, etc. Portfolio assessment, which is more analytic and systematic than descriptive measures, is another means for measuring student outcomes. Another thing that is different from Stake's original model is that in the original model, the evaluator needs to select observers and judges, but in this refined version, the researcher is also the observer. To avoid intra-rater reliability, the

selected classes will be observed over a period of time. Moreover, no judge is required as portrayals are not used in this study.

The researcher's own initiatives are identifications of the types of data needed in this study: context, implementation and student outcomes derived from underpinning concepts of the two approaches, and uses of self-assessment checklist and portfolios as research tools. Context indicators describe the planning stage of the curriculum, e.g., needs analysis, syllabus design and material development. Implementation indicators describe how the curriculum is implemented e.g., instruction, assessment and evaluation. Student outcome indicators describe measurable outcomes: student performance; mastery; attitude, and unexpected outcome naturally occurring in the program. Student portfolios are utilized, instead of portrayals, as the major instrument for assessing their own learning process and outcomes, as well as their perceptions toward learning and the course. A portfolio is more or less similar to a portrayal in that it is an approach to organizing the information about an individual or a class/program. It can contain projects, assignments, various alternative assessments and/or results from NRTs (norm-referenced tests). It is more practical, but less threatening than a portrayal. Moreover, a portfolio can be used to record student achievements in a particular subject area (Wilde and Sockey, 1995). Portfolio assessment can be regarded as a type of personalized assessment as it involves the evaluation of individual students' unique constructions of knowledge and allows them to choose the content and skills to be assessed. Personalized student assessment can be sensitive to the individual student in a manner analogous to the sensitivity of responsive evaluation to individual programs (Mabry, 2001). Learner autonomy is one of the course objectives that can be examined through a self-assessment checklist. Holec (1987 cited in Tudor, 1996) suggests that good language learners are those who are capable of their own learning, and self-assessment plays a central role in the development of learners' self-directed activities. While curriculum-based assessment is a key measure for assessing students' achievement, portfolio assessment as well as self-assessment are measures of students' abilities in taking control of their own study. In brief, based on the two models, curriculum-based achievement tests, classroom observations, material evaluation and interviews are selected as evaluation tools. However, some information, such as learning process, cannot be measured by using such tools.

Therefore, a couple of instruments, including portfolios and self-assessment checklists, are selected and developed by the researcher of this study.

## **2.6 Research in language program evaluation**

Current views of program evaluation put much emphasis on the need for information on language learning concerning both product and process. The qualitative-quantitative debate has brought about an alternative approach which is a combination of methods from the two paradigms. Consequently, for the past twenty years, research on program evaluation in the field of applied linguistics has been conducted using quantitative or qualitative methods or a combination of both.

### **1. Quantitative research**

Nilrat (1986) carried out her research to determine the effects of the two types of EFL programs, an education EFL program and a humanities EFL program, on students' English skill performances. She has examined the relative English language strengths and weaknesses of EFL majors in the faculties of education and humanities at three Srinakarinwirot University campuses in Thailand. A TOEFL test was administered to the subjects of the study and four elements of the TOEFL score were analysed. The findings showed that the two EFL programs at SWU did not have significant differences in listening comprehension, structure, or general English skills, but they did have a significant difference in reading comprehension skills. Both groups scored the highest in language structure and the lowest in listening comprehension. The TOEFL scores of education EFL majors were lower than those of humanities EFL majors.

Ten years later, Chou (1996) evaluated instructors and courses at the Southern Illinois University at Carbondale (SIUC) through the use of Instructor and Course Evaluation (ICE) forms administered to students at the end of each semester. This research examined the relationships between males and females, transfer and non-transfer students, class levels, expected grades and outside-study hours per week. An Analysis of the variance for the results of evaluation between students in required and non-required courses were also provided. The results showed that the male students rated the overall instruction slightly higher than did the females. However, in the overall evaluation, students were satisfied with their education at SIUC.

In addition, Wang (1999) studied the effects of a modified English listening comprehension curriculum on students' achievement in English listening

comprehension at Tamsui Oxford University College in Taiwan. As an experimental research, the experimental and control groups were treated with either the modified or the traditional curriculum. The instruments included student questionnaires, a Michigan Listening Comprehension posttest, and a final examination. Frequency distributions; Chi-square tests, t-tests, and analyses of covariance were used to analyze the data. The findings indicated that there were no significant differences in the students' attitudes and interest in English listening comprehension between those who were taught with an English listening modified curriculum and the control group taught using the traditional curriculum. There was a difference in the final examination scores with the control group scoring higher than the experimental group. However, there was no difference in scores on the Michigan Listening Comprehension Test (posttest). Also, it was found that learning attitude and motivation influenced learning.

One year later, Montañano (2000) investigated the effectiveness of the ENGLABAS program of the English Language Department through the students' pre-test/post-test results. Two parallel tests were research tools, and the pre-test/post-test control group design were utilized to find out whether there was a significant gain between pre-test/post-test scores of the ENGLABAS students compared to the ENGLONE students, i.e. those in the regular English One program. Results showed that the pre-test/post-test mean gain score of the ENGLBAS students was very much higher than that of the ENGLONE students. The ENGLBAS program was successful due to its built-in features, such as added classroom contact hours and small class size.

## 2. Qualitative research

Ward (1987) studied the impact of the Bellon and Handler curriculum evaluation framework of the English as a Foreign Language program at the Huffco Indonesia intensive English Language program in Balikpapan, Indonesia. This study was carried out to provide EFL program administrators with an efficient, effective, systematic, and comprehensive guide to curriculum evaluation. The curriculum evaluation framework, developed by Dr. Jerry Bellon and Dr. Janet Handler of the University of Tennessee, was applied to an oversea ELT program, and its impact was also analyzed. To collect data, the researcher interviewed teachers and administrative staff, reviewed written documents of the Manpower Training and Development Department of Huffco, Indonesia, the department in charge of the intensive English program, and recorded the comments and actions of all respondents outside of the



interview setting. The information on the respondents' reaction during the interviews was also recorded to answer the questions concerning problems emerging during the evaluation and ways to avoid and/or solve these problems, positive results of the education process, and unexpected outcomes.

In the following decade, Alfallaj (1998) investigated the current English curriculum at the College of Technology, Buraydah, Saudi Arabia (CTB). The aims of this study were to evaluate four aspects: (1) the needs of CTB students and their potential employers; (2) the goals of both students and the CTB behind the English course; (3) the nature and effectiveness of materials and media; and the main characteristics of teaching English at the CTB. As a qualitative study, questionnaires, interviews, negotiations, observations, opinion gathering, and existing information were used as research tools. The participants involved in this study were students, teaching assistants, employees who graduated from the CTB, English teachers, teachers from the six departments, representatives of ten local companies, and administrators. The findings revealed that: (1) The current English curriculum was not based on the needs of the students and the companies that hired them, nor was it based on the goals of the students or those of the CTB. (2) The materials and media used at the CTB (limited to the listening lab) were not suited to the students. (3) The major characteristics of teaching English at the CTB were the exclusive use of Arabic, the passiveness of the students, lack of conversation or communication in English, too much grammar, memorization of compositions and grammatical rules, little use of audio-visual media, and teacher-centered classrooms.

### 3. Research using both quantitative and qualitative methods

Chen-Wang (1996) conducted a formative evaluation of the English Language Program in Fong-Sin Senoir High School, Kaohsiung County, Taiwan to examine the effectiveness of the nationally mandated curriculum of the senior high school English Language Program in Taiwan. Stake's (1967) model was used to detect possible discrepancies among nationally mandated curriculum goals, classroom transactions, the alignment between what has been taught and what has been tested in school. Data collected included teacher and student interviews, video-taped classroom observations, translations of the curriculum guidelines, and test scores. Findings showed that even though the nationally mandated senior high school EFL standards expected students to be trained in listening, speaking, reading, and writing skills equally, teaching activities were dominated by grammar translation techniques to train



students with reading skills and small amounts of student-centered writing activities. Chinese was the main instructional medium. Analysis of test scores revealed a relatively high correlation between school tests and the Joint College Entrance Examination. However, both kinds of testing only tested students on reading and writing skills.

Two years later, Bee Bee (1998) evaluated an English teaching project called the Communicative Teaching of English to Medical Personnel (CTEMP) in Minority Autonomous Region in China. The three-week CTEMP project used a communicative teaching approach to help students acquire listening and speaking skills. Feedback was collected from the persons involved: the students and the authorities of the institutions participating in the project. In collecting such data, Bee Bee used different methods, such as a needs analysis, a questionnaire, interviews, and classroom observations. The students' learning outcomes were also assessed through a pre-test and a post-test. This study reported an unanimous agreement on the advantages of a communicative method of teaching that provides the students with more opportunities to use English language. It also helps students to acquire understanding of the use of language in context and increases their motivation to learn. However, this study did not prove whether the students have actually acquired listening and speaking skills during such a short-term project.

In addition, Griffiee (1999) documented a course evaluation of five sections of an Academic English course taught during the spring semester of 1998 at a private university in Japan. Two approaches were employed: quantitative methods and qualitative methods. The quantitative methods included a questionnaire to measure students' confidence in speaking English as a foreign language, criterion-referenced test (CRT) to measure increases in listening ability, and a close-ended (CE) student evaluation of program goals and objectives questionnaire. The qualitative methods included a student open-ended (OE) evaluation questionnaire, multiple student interviews, and multiple teacher interviews. Results of the quantitative measures showed weak improvement in speaking confidence, strong improvement in listening ability, and general student approval of course goals. Results of the qualitative measures showed that the open-ended evaluation questionnaire tended to confirm the results of the close-ended questionnaire (general approval of class activities). The student interviews supported the results of the CRT and the CE questionnaire, but also raised issues not covered by the qualitative instruments. The teacher interviews

revealed that teachers had diverse opinions. For example, their attitude toward the textbook varied between wanting the whole course to revolve around the textbook to rejecting it.

Even though Nilrat's (1986) study was conducted for almost two decades, its content, which is relevant to the EFL proficiency of Thai students and the standards of EFL programs in Thailand, is still very interesting. This study used only a quantitative method in collecting and analysing the data using two control variables: students' GPA's in English courses and the number of years spent studying English. There should be more information concerning the individual student's personal background and experiences that might affect their EFL performance.

The only method used in Chou (1996)'s study was an Instructor and Course Evaluation (ICE) form. The researcher's analyses of 480,845 forms administered to the whole population were time-consuming and boring. According to Wang's (1999) research, there was a difference in the students' achievements in the final examination. The control group taught using the traditional curriculum scored higher than the experimental group taught using the modified curriculum in English listening comprehension, but there was no difference in scores on the Michigan Listening Comprehension posttest. The findings implied that the researcher had to gain more insight into some extraneous variables that would affect the students' performance, for example, teaching methods, students' backgrounds and proficiency levels, timings of test administration, classroom activities, etc. Montañano(2000) did not reveal any evidence of any other types of research instruments except for parallel tests. The claim of the ENGLABAS success might derive from his/her own interpretation of the test scores that could not be accounted as an acceptable or reliable interpretation. A survey of the perceptions of both students and teachers involved should be done to obtain 'real' data that would lead to more insight into these people's attitudes towards the program.

Additionally, Ward (1987) analysed the impact of the curriculum evaluation framework on an ELT program. Interviews were the main research instruments used to collect information concerning the program's status and goals, organization, (operation) curriculum and instruction, outcomes, and recommendations to improve the program. In terms of learning outcomes, with the use of qualitative methods, this research examined only the expected outcomes, not the students' achievements. Alfalaj (1998) used a variety of qualitative methods in collecting the data in order to

gain insights into detailed information on the development and the implementation of the CTB to examine its effects. The participants were representatives of all the stakeholders, the 'real' users of the program, who could provide very useful recommendations for curriculum further improvement.

However, a number of evaluators prefer to combine positivistic and naturalistic approaches in a single study in order to investigate program product as well as process. Chen-Wang (1996) utilized different types of instruments to foster the reliability of this research. However, the selected course did not meet the national curriculum standards. In Taiwan, the nationally mandated senior high school curriculum expects students to be equally trained with listening, speaking, reading and writing skills. It seemed that the main focuses of this course were practices on reading and writing skills, but the tests examined only reading and writing. Bee Bee 's (1998) claim of the use of a triangulation method of evaluation seemed to be very interesting and impressive. Instead of using any statistical device, the researcher utilized only descriptive methods in analysing both qualitative and qualitative data that might devalue the reliability and validity of the research.

Also, in Griffee 's (1999) research, the use of criterion-referenced tests, and student questionnaires as pretests and posttests can be categorized as a kind of summative evaluation. However, it seemed that the researcher tried to utilize the so called 'triangulation method' by using both quantitative and qualitative methods to enhance the reliability of his/her research.

In conclusion, curriculum or program evaluation plays an important role in applied linguistic research. It is the most practical research area that provides fruitful information for the development of the quality of teaching and learning in language education. Second language programs have been developing for more than 30 years, and most of the research in this field is concerned with the evaluation of different types teaching methods, despite knowing that no single teaching method is the 'best'. Similarly, no single evaluation mode would fit all types of curriculum. Mixing positivistic and naturalistic evaluation designs, along with quantitative and qualitative methods, may be the most appropriate design.

### **Chapter Summary**

This chapter provides a theoretical base for the research which involves overviews of task-based approach to language learning, self-directed learning and learner autonomy, and curriculum or program evaluation. The theoretical framework

of the Integrated Stake-Tyler Model, a proposed evaluation model specifically designed for this study is also presented. In addition, this chapter reviews research on program evaluation using quantitative, qualitative and mixed methods that have been done in the last twenty years. The procedure for investigating the effectiveness of the proposed model in evaluating a task-based English course is described in Chapter III.