Pan-Pacific Association of Applied Linguistics 16(1), 123-144

Preliminary Aspects of Language Course Evaluation

Mohammad Zohrabi
The University of Tabriz

Zohrabi, M. (2012). Preliminary aspects of language course evaluation. *Journal of Pan-Pacific Association of Applied Linguistics, 16*(1), 123-144.

Program and /or course evaluation is a process in which different types of data are collected systematically in order to study the virtues and weaknesses of a language instruction program. Program evaluation is, in fact, one of the essential aspects of any curriculum. It is a kind of quality control in which various aspects of an instructional program or course are explored. Program evaluation is an attempt in which different elements of a given curriculum are scrutinized in depth. To this end, an evaluator makes every effort to collect information from different sources such as students, teachers, administrators, course designers, program staff, and so on. Also, the evaluator tries to gather data through different instruments such as field notes, questionnaires, interviews, observations, and course documentations. Collecting hybrid type of information helps increase validity and reliability of the study. The main concern of program evaluation is to ensure that acquisition is taking place, teaching techniques and strategies are useful, materials are relevant and interesting, resources are available and adequate. Finally, the end product of an evaluation endeavor is the improvement and modification of a course of study. This article tries to elaborate on the preliminary aspects of program evaluation.

Key Words: course evaluation, evaluation steps, types of data

1 Introduction

There are various ways in which evaluators can begin their exploration. Before implementing the actual process of evaluation, evaluators should know why they are conducting it. Therefore, determining the objectives helps the evaluators to be specific in their investigation. Also, the evaluators ought to clarify the time in which the evaluation process might be carried out. That is, whether it should be done at the beginning, middle or end of the instruction. Moreover, the elements that are to be scrutinized should be determined before the actual process of evaluation takes place. The crucial issue, at this juncture, is the selection of an appropriate and knowledgeable expert who might carry out the evaluation. Therefore, before the evaluation

process begins, it should be clarified whether an insider, outsider or both of them take this responsibility and perform the evaluation. Furthermore, it should be made clear for whom the evaluation is taking place. In the beginning, program evaluation was merely quantitative and/or experimental in nature. However, with the passage of time it was realized that qualitative and/or natural approaches are more useful than laboratory-like artificial experimental methods. The important issue in any course evaluation is to bring about some necessary changes.

Program and/or course evaluation is a process in which the functioning of a language instruction is explored from different perspectives. Mainly, program evaluation is an essential element in any curriculum. It can be implemented at different stages. Program evaluation is not a product to be studied at a specific point in time. It is an ongoing process which has several key steps. In order to obtain some optimal results, the evaluator should consider these steps carefully. To this end, several researchers have presented different steps for program evaluation, for instance, Parlett and Hamilton (1976), Beretta (1996), Alderson (1996), Nunan (1999), Brown (1989) and Lynch (1996a & b). However, Lynch's (1996b) *context-adaptive model* is rather detailed and complete. This article attempts to delineate different researchers' proposed steps and elaborate on Lynch's model at length. As Lynch notifies, his model is flexible and can be adopted and used in any program evaluation process.

2 History of Program Evaluation

It goes without saying that the most influential scientist in the field up till now has been Ralph Tyler (1949). In Tyler's approach the predetermined issues were compared with the existing issues. At first, behavioral goals were identified, and then tests were developed based on those goals. This approach had some defects. For instance, there were some unpredictable issues that were abstract and could hardly be determined and defined. Therefore, limiting evaluation to only behavioral objectives left out these issues. Also, one of the main deficiencies of Tyler's approach was its lack of attention to process. It practically disregarded what actually took place within an instructional program.

Obviously, in Tyler's approach there is no room for qualitative and/or non-quantifiable information. There is also no consideration being paid to the actual process of language learning and teaching. The main factors in this approach are the measurable products and behaviors. In fact, this approach only assesses the students' attainment of course objectives. Also, the information gathered at the end of the program cannot be fed into the course and has rarely any feedback usefulness. Moreover, there is hardly any data collected on the perceptions of students, teachers and administrators. Therefore, product-oriented approaches have relatively many shortcomings

which little by little have lost their credibility in the last decades. It can be stated that this approach can hardly be considered as an appropriate way of implementing the program evaluation. This approach is merely testing the students' attainment of course objectives which can be usually fulfilled by any language teacher. The following two studies represent the traditional approaches to program evaluation.

The first study is concerned with Raymond F. Keating's (1963) experiment on the effectiveness of language laboratory use. Keating tested two groups of students: the *experimental group* was taught through laboratory but the *control group* received regular treatment. However, both groups were assessed by means of reading, listening and speaking tests. Nonetheless, the control group students outperformed the laboratory or the experimental students. Nevertheless, this study was criticized for its lack of validity, lack of controlling various variables, and faults in its experimental design.

The second entirely controlled experimental method study was conducted by Chastain and Woerdehoff (1968). They too compared the efficacy of audio-lingual and traditional methods. They referred to the traditional approach in their study as the cognitive code method. In order to control different variables, they administered a pretest at the beginning of the course and a posttest at the end. The researchers concluded that the cognitive code method was much better than the audio-lingual method.

Generally, one evaluation model which was in vogue in the 1970s was Stufflebeam et al.'s (1971) CIPP evaluation (Context, Input, Process, and Product). In this model, *context* evaluation is concerned with the evaluation of a program of study to explore its strengths and weaknesses in order to make it better. *Input* evaluation determines the available means through which to evaluate the objectives of a program. In input evaluation the evaluator attempts to use every resource to carry out the evaluation. *Process* evaluation is the actual phase of fulfilling an evaluation. The evaluator employs several techniques in order to execute the evaluation and in this way works toward the improvement of the program. In the end, the *product* evaluation tries to assess how far the program objectives have been achieved. On the whole, this model attempts to supply information for decision-makers.

A nightmare for an evaluator is to come to terms on what to evaluate. However, it would be better to choose the factors that have either direct or indirect effect on the students' progress rate. That being so, "it is up to individual teachers and curriculum personnel to decide how widely they should cast the net" (Nunan, 1999, p. 119). Nunan also believes that the likely factors for evaluation might include "initial planning procedures, goals, content, materials and learning activities, teacher performance and the assessment processes, and learner achievement" (ibid.). Yet, Brown (1989, p. 283) suggests that different program components can be evaluated from different points of view:

Figure 1. Evaluation components and viewpoints

Viewpoints	Needs	Objectives	Testing	Materials	Teaching
Effective?					
Efficient?					
Attitudes?					

3 Different Proposed Steps

In fact, stages of program evaluation are one of the critical phases of planning the evaluation process. As Alderson and Beretta (1996, p. 273) remind, it is at this phase that "the brief of the evaluation needs to be carefully and fully discussed…" The first phase of deciding on the appropriate ways of executing evaluation is important because it can hardly be rearranged later. There are several consecutive steps in conducting the evaluation process. To this end, several different forms of implementing program evaluation have been proposed by different researchers. An evaluator can choose one of the proposed forms and adapt it according to his/her own and program's criteria. At this juncture, Alderson (1996, p. 274-5) states that the choice to be made "depends on the purposes of the evaluation, the nature of the program …, the individuals involved … and on the time scales and resources involved." With regard to these issues, the evaluator ought to attempt to conduct the evaluation process in as much clear and systematic way as possible.

Therefore, in order to achieve this goal, four standards of evaluation have been proposed by Beretta (1996, p. 18): "the utility, feasibility, propriety and accuracy standards." The *utility* standards are concerned with the responsibility of the evaluator to provide the stakeholders with appropriate information and data. The *feasibility* standards demand that the evaluation design be practicable in particular contexts. The *propriety* standards require that the evaluators act properly and consider the rights of those who might be influenced by the research process. Finally, *accuracy* standards demand that the evaluation process be well-grounded and the information be sufficient. As it was noted earlier, different proposals for different stages of evaluation have been put forward by different researchers. For instance, Parlett and Hamilton (1976) identify three stages for implementing evaluation:

1- Observation
2- Further enquiry and narrowing down the information
3- Description and explanation of the findings.

The above mentioned three stages are too broad and, therefore, some specific and detailed stages are needed. Beretta (1996) also offers three stages

of evaluation but does not elaborate on them as fully as possible. Therefore, his stages can be summarized as follows:

1- Negotiation periods: In this stage the evaluator and the stakeholders attempt to reach an agreement on the principles of evaluation.
2- In the second stage the evaluator tries to find an appropriate design, collect the data and analyze them.
3- Finally, the evaluator seeks the ways of putting findings into action and bringing about change in the program.

Meanwhile, one more proposal has been made by Alderson (1996, p. 274). In his outline there is no mention of actual ways of implementing the evaluation process. He has merely listed some stages without giving any appropriate explanations about them:

1- Planning: Purpose: Why?
2- Implementing: Audience: Who for?
3- Interpreting: The evaluator: Who?
4- Reporting: Content: What?
5- Using: Method: How?
6- Evaluating: Timing: When?

On the other hand, Nunan (1996, p. 201) sets forth an overall picture of the evaluation process. Again his description is broad and there is hardly any clarification of the actual stages of executing evaluation:

- Purpose: What is the purpose of evaluation?
- Audience: Who is the audience for the evaluation?
- Principles of procedures: What principles of procedures should guide the evaluation?
-Techniques and instruments: What tools, techniques and instruments are appropriate?
- Data analysis: Will the data analysis be statistical or interpretative?
- Evaluation: Insider/outsider: Who should carry out the evaluation?
- Time frame and budget: What is the time frame and budget for the evaluation?
- Reporting: How should the evaluation be reported?

Moreover, Brown (1989, p. 222-41) advances six stages of program evaluation as follows:

1- Creating a conceptual framework
2- Determining the theoretical foci
3- Formulating research questions
4- Selecting Procedures
5- Data gathering process
6- Data analysis and synthesis of information.

Generally, Brown's stages sound rather appropriate; however, there are some shortcomings too. For example, the purpose or the objective of the evaluation is missing. Also, stage two is vague and ill-defined. Stage four has been rather misplaced because it should be located at the first stage. Moreover, there is a grave overlap between stages four and five. Both of these stages are relatively similar and can be merged into one stage. Finally, reporting of the findings is missing.

Finally, Lynch (1996b, p. 3) puts forward a fairly elaborate and comprehensive format for program evaluation. He refers to his model as context-adaptive model (CAM) because it can be adapted in evaluating any instructional program. The steps of program evaluation are as follows:

1- Audience and goals
2- Context inventory
3- Preliminary thematic framework report.
4- Data collection design/system
5- Data collection
6- Data analysis
7- Evaluation

3.1 Participants and objectives

Generally, the first stage of an evaluation is very important because it is at this stage which the audience and their goals of conducting evaluation are determined. In every evaluation there are different audiences with different purposes. To clarify who the audience are, Lynch (1996b, p. 3) raises two key questions: "Who is requesting the evaluation? Who will be affected by the evaluation?" The first question might be put forward by the funding agencies who want to see value for money. These stakeholders want to know whether or not the students' test scores are high enough and satisfactory in order to continue supporting the program financially. The second stakeholder might be the institution itself which may want to see whether or not the program is successful. The second question is related to the students and to some part to the teachers who might be affected by the results of the evaluation. Therefore, the audiences of an evaluation might include funding agencies, educational institutions, administrators, program staff, curriculum developers, parents, teachers, other researchers and evaluators, and students who might be interested in the evaluation results.

To this end, to determine the goal(s) of an evaluation, Lynch (1996b, p. 3) sets forth two questions: "Why is the evaluation being conducted? What information is being requested and why?" It is clear that different stakeholders have different goals. After the audiences of the evaluation have been determined, the evaluator attempts to reach an agreement among them in order to conduct the evaluation. At this stage, there are various points of view and, of course, expectations. Obviously, the evaluator cannot carry out the evaluation process according to his/her own objectives. As Alderson (1996) points out, every stakeholder has his/her own goals and even these goals are sometimes in conflict with each other. Mainly, the evaluator should clarify the specific audiences and their particular aims. Therefore, before conducting an evaluation, the evaluator should try to find out answers to these questions: "Why am I doing this research? In whose interests is this research?" (Blaxter et al., 2006, p. 10-13). However, Brown (1989 & 1995) considers goals of fulfilling an evaluation from four different points of view.

It is certain that an evaluation can be approached from different perspectives based on a program's goals and audiences. Generally, there are several factors which influence a choice of one approach over another. Mainly, the key elements which affect our choice include the audience, objectives and goals of the research problem, to name but a few. Therefore, according to a program's particular goals, Brown (1989, p. 224 & 1995, p. 219-20) offers four approaches to program evaluation. Therefore, any program evaluation might be carried out based on one of the following approaches:

1- Goal-attainment and/or product-oriented approaches
2- Static-characteristic approaches
3- Process-oriented approaches
4- Decision-facilitation approaches.

3.2 Developing a checklist

Mainly, the second step in the process of program evaluation is the context inventory (Lynch, 1996b). In order to evaluate a program, the evaluator should make a checklist to identify the features that should be considered and acted upon. These dimensions of the program should be identified at this stage in order to guide the evaluation processes. To this end, Murphy (1985, p. 1-17) states that an evaluation might look like as follows:

- Scope
  - curriculum
  - materials
  - teaching & learning
  - attitudes

- Purpose
  - to relate elements of curriculum to each other and to goals and effects
  - assessment
  - accountability
  - awareness

- Methods
  - observation & description
  - qualitative & quantitative methods
  - subjective & objective techniques

Furthermore, Alderson and Scott (1996, p. 43) offer eleven categories in the design of the program evaluation. They believe that these categories help to gather data and analyze them. These categories are as follows:

Context

1- Attitudes and motivation with regard to the project

Methodology

2- The ESP methodology or approach adopted

3- Classroom management

Implementation of methodology

4- Materials

Project achievements

5- Learning outcomes (students' learning)

6- Impact on 'outsiders'

Teacher-training implementation

7- In-service training

8- Publications

9- Research carried out by teachers

Exchange of ideas and experience

10- The administration of the project

11- Resource center.

On the other hand, Lynch (1996b, p. 5) suggests a fairly comprehensive checklist for program evaluation as follows:

- Availability of various types of evaluation expertise.
- Timing of the evaluation.
- The selection process for admitting students into the program.
- Characteristics of the program staff.
- Size and intensity of the program.
- Instructional materials and resources available to the program.
- Perspective and purpose of the program.
- Social and political climate surrounding the program.

Mainly, the flexible nature of the CAM (context inventory model) allows the evaluator to choose from among the various program factors and tailor them according to the specific program that should be evaluated. The program evaluation needs to consider these dimensions in its evaluation process based on their particular setting. This checklist can inform the evaluator of the limits of the peculiar program that he/she intends to evaluate.

3.3 Initial framework

It is after determining the related audiences and their particular goals that the evaluator should attempt to develop a framework. This framework is of paramount importance because it shapes the whole process of what should be done. Creswell (1994, p. 97) emphasizes the significance of an early framework and declares that "the researcher advances a tentative conceptual framework in a qualitative study early in the discussion." Generally, the frameworks can appear in different forms and sizes. They can come in simple or detailed shapes; also they might be explanatory or in causal relationships, based on a theory or practicality of use. To this end, Miles and Huberman (1984, p. 28) contend that "A conceptual framework explains, either graphically or in narrative form, the main dimensions to be studied – the key factors, or variables – and the presumed relationships among them." By and large, a framework creates a working outline in order to focus the evaluation process. It indicates the direction of the evaluation and the elements that might be evaluated. Lynch (1996b, p. 6) confirms that it "provides a conceptualization of the program in terms of the salient issues and themes … [and] provides the evaluator with a focus that will guide the collection and analysis of evaluation data." In order to devise a framework based on evaluation objectives and design of the study, Miles and Huberman (1984, p. 33) suggest that:

- Conceptual frameworks are best done graphically, rather than in text.
- One should expect to do several iterations of these frameworks.
- Develop simplified frameworks without arrows going in all directions.

Generally, in order to carry out an actual program evaluation, Deyes (1988 cited in Lynch 1996a, p. 291) puts forward a framework which appears in an outline form:

Table 1. An Outline of a Program Evaluation Framework

	Project narrative [or structure]	Indicators of achievements	Means of verification [quantification]	Important assumptions
-Wider objectives -Immediate objectives - Output - Input				

Deyes's framework is in the form of an outline rather than a drawing. The evaluator puts information into the blank spaces during the evaluation process. This outline provides an overall picture of what is being done. However, what is needed is an actual figure which could guide the evaluator during the different phases of the evaluation process. That picture could be of great help in shaping the evaluation process and delineating the objectives, elements, and different steps that might be taken.

Obviously, the conceptual framework should be relatively comprehensive and depict the beginning and ending process of the evaluation process. It should illuminate every minute detail and the steps that need to be taken in the process. The framework should be cyclical in nature and ongoing. That is, the evaluator might return to the previous stages and reassess the work again and again. Also, every phase of the evaluation should be ongoing and dynamic. In the main, the framework might consist of different steps which all together can render a whole picture.

3.4 Research design and data type

The evaluator should make every effort to choose a robust and adequate design based on the program's various stakeholders and his/her own objectives, so that to evaluate a program properly, To this end, an evaluator can opt either for a quantitative, qualitative or a mixed method design. Generally, Lynch (1996b) refers to quantitative designs as positivistic research and to qualitative designs as naturalistic research.

3.4.1 Quantitative research

In fact, quantitative research is called traditional approach because it was dominant in the 1960s and 1970s. In general, Richards and Schmidt (2002: 436) state that quantitative research "uses procedures that gather data in numerical form … [It] aims at causal explanation of phenomena through the identification of variables which can be made the basis of experimental investigation." In most cases the focus of attention is on the end product, i.e. the students' attainment of course objectives. Also, the program evaluations

merely consist of comparison of different teaching methods. Mainly, students are divided into two groups: control and experimental groups. The experimental group receives a treatment and the control group is taught through the regular program. A pretest and posttest is administered and the students' scores are studied accordingly. Alderson (1996, p. 283) confirms that "the results of such experimental methodologies were less than encouraging, and … such approaches are inappropriate to program evaluation." In this regard, Long (1984) also disapproves of laboratory-like, experimental methods because of their lack of attention to the processes of teaching-learning. That is, without the proper description of the process the product will be meaningless. At this juncture, Lynch (1996b, p. 96) argues that "it is not enough simply show that a program was successful on some outcome measure. It must also be possible to make some judgments of what … made the program successful." In the main, there are some differences between quantitative and qualitative researches.

It goes without saying that quantitative and/or positivistic research can hardly be of any use in program evaluation. In a program evaluation there are various unexpected occurrences which the laboratory-like experimental designs cannot account for them. However, in the qualitative and/or naturalistic researches the evaluator takes part in the actual process of evaluation and is always involved in the program rather than being an unknowledgeable outsider. In naturalistic approaches the evaluator usually does not disrupt the normal processes of the classroom activities. Generally the qualitative approach is exploratory in nature and does not attempt to verify predetermined hypotheses. Finally, qualitative approaches presume that a program is a dynamic reality which changes over the time and is not stable.

3.4.2 Qualitative research

Qualitative research allows the evaluator to investigate the research setting, context, and individuals' activities closely. This type of research attempts to study the participants and their performances in their natural milieu. For this reason, it is also referred to as naturalistic research. Therefore, naturalistic research is heuristic with barely any predetermined questions or hypotheses to dictate the research process (Seliger & Shohamy, 1989). Seliger and Shohamy also contend that this approach is very useful when an evaluator attempts to study the learning-teaching process in its natural setting. Blaxter et al. (2006, p. 64) elaborate on qualitative research as follows:

> Qualitative research … is concerned with collecting and analyzing information in as many forms, chiefly non-numerical, as possible. It tends to focus on exploring, in as much detail as possible, small numbers of instances or examples which are seen

> as being interesting or illuminating, and aims to achieve depth rather than breath.

Generally, in this type of inquiry the evaluator tries to focus on the specific context in which the evaluation is being carried out and at the same time "collaborates with the participants" (Creswell, 2003, p. 19). Qualitative research is innovative as well as emergent and is concerned with individuals and their perceptions rather than with numbers and figures which are abstract and out of context. Lynch (1996b, p. 14) insists that naturalistic paradigm "views the educational program being evaluated … as a process that is continuously changing rather than a stable, invariant treatment." Lynch also emphasizes that as more evaluators use this approach in the field of English language teaching, therefore, more sophisticated and elaborate ways of adapting its methodology to a particular context can be seen. By and large, naturalistic approaches allow the evaluator to collect data through different techniques such as observations, interviews, questionnaires etc. and also from various sources such as students, teachers, administrators and so on. Consequently the variety of sources and techniques makes the data more and more valid and convincing.

Thus, the numerous advantages and suitability of the naturalistic approach renders it as one of the main candidates of program evaluation. Some studies, however, might use a mixed method and/or mixed strategy (e.g. both qualitative and quantitative) as their research design. Therefore, the model which best represents a mixed method design is the illuminative model which is discussed in the next section. Although illuminative model is one of the branches of naturalistic inquiry, it uses both qualitative and quantitative data. Therefore, Lynch (1996b, p. 82-4) strongly recommends it as one of the ideal designs for program evaluation.

3.4.3 The illuminative model

On the whole, the primary goal of a study might be to describe the various aspects of a language instruction course. As such the main emphasis is on the process of what is happening inside and outside the classroom. So that the rather preferable design for this inquiry might be Parlett and Hamilton's (1976) illuminative model. This model makes room for both qualitative and quantitative data collection. The emphasis of this model is on investigating, observing, and describing the situation under study. Lynch (1996b, p. 84) reasons that since illuminative model makes use of both qualitative and quantitative data and analysis "It is, perhaps, better thought of as a mixed strategy." The most significant characteristic of illuminative model is its observation aspect. That is, the evaluator, at first, attempts to understand the problem(s) under study through careful observation. This close observation

of the events helps the researcher gain an understanding of what is happening and the nature of the issues under study.

In the last decades, there have been a wide range of program evaluation models available for the evaluators to select from and conduct their evaluation. One of the most prominent program evaluation models within the naturalistic design is the *illuminative model* developed by Parlett and Hamilton (1976). One of the significant advantages of this model is its utilization of both qualitative and quantitative data and analysis. For this reason Lynch (1996b, p. 83) vehemently affirms that "illumination model is a mixed design rather than a naturalistic one." He also argues that this model is at times called as transactional as a result of "its focus on multiple audience perspectives and program process" (ibid. 82). Parlett and Hamilton (1976, p. 144) clarify their model's aims as follows:

> The aims of illuminative evaluation are to study the innovatory program: how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected.

It can be argued that illuminative model might be considered as one of the good examples of a nontraditional procedure for exploring the success or failure of a program. Richards (1984) elucidates illuminative model as a custom built research strategy which generally keeps away from (though not ruling out) statistical procedures. This model attempts to obtain a wide variety of data on the program and its milieu and consequently work out its complexities as far as possible. This model's advocates emphasize the importance of the process over the product (Ramsay & Clark, 1990). Also, one salient feature of the illuminative model is that it attempts to investigate all the aspects of the program and find out answers to different questions that may arise during the evaluation process. Richards and Schmidt (2002, p. 247) describe this model as:

> an approach to evaluation that seeks to find out how different aspects of a course work or how a course is being implemented and the teaching-learning and processes that it creates. It seeks to provide a deeper understanding of the processes of teaching and learning that occur in a program …

There are usually three stages in which illuminative model can be implemented:

1- Observation,

2- Further inquiry and narrowing down the information,
3- Description and explanation of the findings.

In the first phase, the evaluator tries to obtain a holistic picture of the program's overall workings. He/She attempts to become acquainted with every aspect of the program and its context. After this initial period of observation, the evaluator makes every effort to derive minute issues and themes in order to narrow his/her studies and focus on the details. Parlett and Hamilton (1976) consider the movement from more general events to detailed ones as the progress focusing. Finally, the evaluator combines the findings and explains them to the intended audience. One of the salient features of this model is that it deals with the unexpected happenings which may surface during the evaluation process. However, there is hardly any rigid borderline between these stages. "In practice, these stages overlap both temporally and functionally, and go on and on until, after successive inquiries, a clear focus is obtained" (Beretta, 1996, p. 17).

There are several major data collection techniques within the illuminative model: observation, interview, questionnaire, program documentation and so forth. By and large, observation is of prime importance in the initial stages of the evaluation. Then, the evaluator interviews program administrators, teachers and students. In order to increase the validity of the data, the evaluator gathers more information through questionnaires which are filled out by students, teachers and administrators. Moreover, the program's history, development and objectives can be investigated through program documentations.

Finally, the illuminative model has received some criticism now and then. For instance, Crittenden (1978 cited in Ramsey & Clark, 1990, p. 38) claims that:

> First, by stressing the uniqueness of each setting, illuminative models do not produce findings which have any generalizability. Second, relying on the perceptions of the observer introduces problems of subjectivity. Third, the desirability, or even possibility of the evaluator remaining judgment free is questionable.

These criticisms are barely acceptable because illuminative model has many advantages which make it almost one of the best methods of evaluating a program. First, it is an accepted fact that every educational setting has its peculiar characteristics. However, there are also many similarities which cannot be ignored so easily. Furthermore, the aim of conducting an evaluation within a particular context is to find out its strengths and weaknesses. Therefore, the evaluation results are used to tackle that specific setting's problems and not to generalize them, though some generalizations

can also be made to similar settings. Second, it is a myth that the findings are acquired merely as a result of the evaluator's perceptions. On the contrary, most of the findings are obtained based on program's stakeholders such as students, teachers and administrators through interviews and questionnaires. Finally, though the ultimate decisions are made by administrators and program organizers, it is the evaluator who observes, narrows the study and synthesizes the findings. It is, in fact, the evaluator who combines the various data and prepares the final report. On the whole, the illuminative model has many benefits which make it one of the prominent forms of program evaluation.

3.5 Data collection

Naturally, the type of data that the evaluator gathers almost entirely depends on the type of design that has been chosen. The fifth step in program evaluation process according to Lynch (1996b, p. 4-8) is the data collection procedure. If the aim of the evaluation is to enquire about the students' achievements of course objectives then the evaluator can gather data through the students' end-of-program tests. On the other hand, if the goal of evaluation is to improve the whole course, then the evaluator can collect data through questionnaires, interviews, observations, course documents and so on. However, the important point that the evaluator should take care of is "the appropriate conduct of the data-gathering procedures" (ibid. 7). Therefore, the evaluator should decide on the pertinent types of data to be collected in order to make necessary interpretations according to them. As Murphy (1985, p. 15) stresses, we should "know what sort of information we want and the appropriate ways to choose from to get it." It can be concluded that the evaluator should try to gather not only relevant data but also through suitable ways.

The evaluator can opt either for qualitative or quantitative data to gather information. As usual, it depends on the purpose for which the evaluation takes place. However, for any evaluation to be rich and have convincing evidence, it is preferable that both types of data to be collected and used. Generally, qualitative data can be described as the type of information which can be obtained through observations, interviews, open-ended questionnaires and so on. Alderson (1996, p. 282) believes that exploring "attitudes and opinions are important to the evaluation." He then emphasizes that "observation, recording and interpretation of events, activities and feeling of participants" (ibid. 283) are also of paramount importance. In addition, Brown (1995, p. 227) assumes that qualitative data contains "more holistic information … that may not readily lend themselves to conversion into quantities or numbers." Nevertheless, Richards and Schmidt (2002, p. 435) assert that qualitative data "can often be converted into quantitative form." On the whole, in order to interpret the qualitative

data, the evaluator can utilize them "in a principled and systematic manner [because] they are more important to actual decisions made in a program" (Brown 1989, p. 232).

In contrast, quantitative data are countable pieces of information which are usually numerical in form. They can be obtained through "tests and objective-question questionnaire" (Dudley-Evans & St John, 2000, p. 128). Qualitative data ate mostly obtained through observations, interviews and open-ended questionnaires and quantitative data are gained through closed-ended questionnaires. A researcher might not use testing procedures to collect data because as Elley (1989, p. 270) affirms "Samples are biased or unmatched, contamination occurs between experimental and control groups, and tests prove too difficult or too easy for students." Also, it can be stated that though tests may be important, "they are not the exclusive, or even the primary, focus of all evaluations" (Nunan, 1999, p. 190). Moreover, the information that can be gathered through tests has rather limited interpretational value because "They provide answers to what questions but cannot easily address the how or why" questions (Dudley-Evans & St John, 2000, p. 128). More importantly, tests assess what teachers think are the objectives of the course, however, many vital and unexpected issues surface incidentally in the actual teaching-learning processes. To this end, Slimani (1996, p. 199) reasons that "However, since we are concerned with relating learning outcomes to their immediate and potentially determining environment, it appears rather difficult to think of ways of getting at learning evidence through testing and elicitation procedures as traditionally understood."

One more issue in data collection process is the diversity of the information that is gathered. It is reasonable that the evaluator attempt to gather as varied sorts of data as possible. Creswell (2003, p. 21) affirms that "collecting diverse types of data best provides an understanding of a research problem." Therefore, collecting more miscellaneous data increases validity and consequently adds to the legitimacy and authenticity of the evaluation endeavor. For instance, Nunan (1999, p. 189) argues that the evaluator should also try to inquire into the "institutional facilities, the prevailing intellectual and emotional climate, relationships between administrative and teaching staff, and so on."

The list suggested by Brown (1989, p. 233) is rather comprehensive and detailed. However, an evaluator cannot gather data through all these procedures because of the constraints of time, money and expertise. Brown (ibid.) himself asserts that "Obviously, it would be absurd to attempt the use of all the procedures … but a reasonable selection can be made…" By and large, the evaluator should attempt to gather data according to the objectives of the evaluation. If the researcher tries to collect more than necessary data, then he/she will, naturally, be lost among the mass of unnecessary information. On the other hand, if he/she gathers less data, he/she will not

gain enough insight towards the problem. A moderate amount of information is usually more preferable.

The important issue that any evaluator might encounter is the conflicting types of data that are gathered. Hutchinson and Waters (1987, p. 155) argue that "feedback from one party [e.g. learners or teachers] may contradict feedback from another." This is a type of problem that usually an outsider evaluator might face because of his/her rather incomplete knowledge on the setting from which data are collected. On the other hand, an insider evaluator may fairly have thorough understanding of the context and tackle the issue more easily. In this regard, the evaluator should try to gather more data and, of course, from different sources in order to gain more insight into the problem and into the contradictory data.

3.6 Data analysis

The evaluator ought to opt for the type of data analysis based on the design that has already been selected and carried out in the data collection process. Therefore, the sixth step in program evaluation as suggested by Lynch (1996b, p. 7) is the data analysis phase which "follows logically from the type of design chosen for the evaluation." However, Creswell (1994, p. 153) claims that "The process of data analysis is eclectic; there is no right way." Nevertheless, this statement does not mean that the evaluator is free to choose any method of analyzing data. It means that within the special design chosen for the evaluation, the evaluator might approach data analysis from different angles according to evaluation objectives, possess necessary expertise in data analysis, and also have necessary facilities and resources at hand.

An experienced researcher goes through the data at several stages and analyses them throughout the whole evaluation process. Generally, data analysis does not take place only at one phase. To this end, Blaxter et al. (2006, p. 193) believe that "Analysis is an ongoing process which may occur throughout your research …" The purpose of this continuous data analysis is mainly to extract meaning from a pile of data. As Creswell (2003, p. 190) contends, it means "moving deeper and deeper into understanding the data, representing the data, and making an interpretation of the larger meaning of the data." In this regard, the evaluator ought to find some appropriate ways of making sense out of a mass of data. Therefore, in order to interpret the data and present the evaluation results, the evaluator should read through the data, analyze, categorize, condense and finally synthesize them (Seliger & Shohamy, 1989). Obviously, data analysis demands that the inquirer form some categories and at the same time make some necessary comparisons and contrasts. The evaluator might face various types of contradictory data during this stage. He/She should make a wise and logical analysis according to the existing evidences. Creswell (1994, p. 153) argues that "the researcher be open to possibilities and see contrary or alternative explanations for the

findings." Therefore, this stage is very important because its effects lead to the results and conclusions of the study.

3.7 Reporting

The evaluator describes the procedures that he/she performed and explains in detail the results or outcomes of the study. Therefore, the outcomes of the whole evaluation process are reported to the relevant audiences. Lynch (1996b, p. 174) states that "The final product of an evaluation is a report, which can take many different forms, depending on the audiences and goals …" As Creswell (1994, p. 169) emphasizes, the important issue is that "how this outcome compares and contrasts with theories and the literature." That is, the results ought to be collated and examined in comparison with other related studies, theories and so on. The evaluator should try to expound the study, the results and his/her interpretations as clearly as possible to the relevant audience. In this way, the readers can make sense of the whole evaluation "and what the study means to them and to the language teaching profession as a whole" (Brown 2001, p. 12).

Sometimes the results of the evaluation may be provocative and create some controversial issues. Therefore, the evaluator must try not to hurt the stakeholders. However, Lynch (1996b, p. 9) believes that "The critical issue is how to communicate the findings of the evaluation honestly and successfully." It is safe to say that the evaluator should consider the social and political atmosphere of the context and situation and act accordingly. This does not mean that he/she might make some compromises. Rather, the evaluator should be considerate and take into account every aspect before reporting the outcomes. Creswell (1994, p. 169) argues that the inquirer should impart his experiences so that "allow readers to vicariously experience the challenges he encounters and provide a lens through which readers can view the subject's world." On the whole, the evaluator should report the results based on the pre-agreed goals and objectives of the evaluation.

The evaluator should report the findings according to the specific design which the data were gathered and analyzed. Generally, there are different formats for reporting the evaluation results. Brown (2001, p. 253) divides the research design into four parts: "purely statistical, statistical with some quantitative, qualitative with some statistics, and purely qualitative." The quantitative research report is usually presented in numerical forms along with the relevant tables, diagrams and figures. However, the qualitative research report appears in narrative form and its organization is fairly flexible. As Creswell (1994, p. 168-9) puts it, "the results will be presented in descriptive, narrative form rather than as a scientific report." The important issue in this process is to represent the outcomes as fully and clearly as possible. In qualitative research the researcher makes every effort to recount the process of research. That is, the process is more important than the sheer

product. As Brown (2001, p. 257) asserts, this account of the "story may differ in structure from project to project and report to report."

Generally, a research could be a mixture of qualitative and quantitative methods. To this end, the organization of the report might be in a hybrid form. Therefore, if the design is a mixed approach, a combination of descriptive and statistical report forms might be rendered. However, Brown (2001, p. 259) cautions that "you may need to decide whether it is primarily a statistical study or mainly qualitative in nature. Therefore, the evaluator should know which of the approaches is the predominant one and accordingly prepare a report on those bases.

4 Conclusion

It is safe to say that program and/or course evaluation is a kind of glue which joins all the elements of a curriculum together. Without program evaluation we cannot make sure whether the students' true needs are met, whether they are satisfied with the course under study, whether the materials are effective, and whether testing motivates more learning or impedes it. Therefore, program evaluation attempts to investigate a course from different perspectives. The important point is that an evaluator cannot properly succeed to carry out an evaluation unless different parties involved in it try to help him/her in one way or another. Program evaluation is, in fact, a vast and broad endeavor. It requires a great amount of time, energy, expertise, experienced personnel, resources and so on. It cannot be performed at a given point in time. It is an ongoing process which begins at the start of the course, continues till its end and even after it. Program and/or course evaluation tries to bring about some necessary and adequate modifications. Therefore, the end result or outcome of an effective evaluation process is to motivate some useful improvements in a course of study.

It is clear that program and/or course evaluation is not just a one-off endeavor; it happens regularly at several continuous stages. The important issue at this process is that the evaluation phases are not separate from one another. That is, they do not take place one at a time, and they are not happening linearly. Rather, they are cyclical and are dependent on each other. Sometimes, there is some overlapping too. However, a meticulous evaluator tries to consider all the steps to determine the audience and their goals, to opt for an appropriate design, collect enough information, analyze data and finally report the findings. The crucial point is that the entire evaluation endeavor should result in some satisfactory and adequate outcomes. The aim of program evaluation is to bring about some effective changes in order to modify and improve a given language instruction course and/or program.

In the main, program and/or course evaluation is a multifaceted area in which various factors and elements should be taken into account. Program evaluation is not a simple one-shot happening. It consists of different

dimensions that determine the way in which varied types of information are collected. Also, program and/or course evaluation is carried out to achieve fairly different goals. Furthermore, program evaluation needs to be implemented through an appropriate and robust design. The design of the evaluation is of outmost importance. It is the design of the study which determines what type of data to be collected and how it should be analyzed and interpreted. Of course, the selection of a proper approach depends on the goals and objectives of the evaluation. In this process, the role of the evaluator is very critical. It is the evaluator who should opt for an adequate research design, collect relevant data, analyze them appropriately and obtain the necessary results. All these efforts are made in order to produce an effective course and hence to modify and improve the learning-teaching process.

Program and/or course evaluation is an extensive area and has different procedures. Before conducting the evaluation process, the evaluator should determine why it should be done. Then, he/she might decide when it should be carried out. After that, the evaluator should clarify the elements that might be investigated. Moreover, the educational institution may determine whether an insider, outsider or both of them should conduct the evaluation task. Furthermore, the different parties who require the conduct of the evaluation process should be identified beforehand. That is, whether it is carried out for the benefit of the students, teachers, administrators, program staff, parents, ministry of education and so on. At this process, the corollary of evaluation process should be determined. That is, the end result of the program evaluation should be clarified. When the evaluator tried to find appropriate answers to different wh-questions, he/she could better approach the problem at hand. Without providing proper answers to those questions, the evaluator cannot do his/her job adequately.

References

Alderson, J. C. (1996). Guidelines for the evaluation of language education. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 274-304). Cambridge: CUP.

Alderson, J. C. & Beretta, A. (Eds.). (1996). *Evaluating second language education*. Second printing. Cambridge: CUP.

Alderson, J. C. & Scott, M. (1996). Insiders, outsiders and participatory evaluation. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 25-60). Cambridge: CUP.

Berreta, A. (1996). Evaluation of language education: An overview. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education*, (pp. 5-24). Cambridge: CUP.

Blaxter, L., Hughes, C. & Tight, M. (2006). *How to research*. Berkshire: Open University Press.

Brown, J. D. (1989). Language program evaluation: A synthesis of existing possibilities. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 222-241). Cambridge: CUP.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston, Mass.: Heinle & Heinle Publishers.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: CUP.

Chastain, K. D., & Woerdehoff, F. J. (1968). A methodological study comparing the audio-lingual habit theory and the cognitive-code theory. *Modern Language Journal*, *52*(5), 268-279.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousands Oaks, CA: Sage Publications.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Second ed. Thousands Oaks, CA: Sage Publications.

Dudley-Evans, T. & St John, M. J. (2000). *Developments in English for specific purposes: A multidisciplinary approach*. Cambridge: CUP.

Elley, W. B. (1989). Tailoring the evaluation to fit the context. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 270-285). Cambridge: CUP.

Huberman, A. M. & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage Publications.

Hutchinson, T. & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge: CUP.

Keating, R. F. (1963). *A study of the effectiveness of language laboratories*. New York: Columbia University.

Long, M. H. (1984). Process and product in ESL program evaluation. *TESOL Quarterly*, *18*(3), 409-425.

Lynch, B. K. (1996a). Evaluating a program inside and out. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 61-96). Cambridge: CUP.

Lynch, B. K. (1996b). *Language program evaluation: Theory and practice*. Cambridge: CUP.

Miles, M. B. & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage Publications.

Murphy, D. F. (1985). Evaluation in language teaching: Assessment, accountability and awareness. In J. C. Alderson (Ed.), *Evaluation* (pp. 1-17). Oxford: Pergamon Press.

Nunan, D. (1999). *Research methods in language learning*. Eighth printing. Cambridge: CUP.

Parlett, M. & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovatory programs. In G. V. Glass (Ed.), *Evaluation studies* (pp. 140-157). Beverly Hills, Calif.: Sage Publications.

Ramsay, W. & Clark, E. E. (1990). *New ideas for effective school improvement: Vision, social capitol evaluation*. Hampshire: The Falmer Press.

Richards, J. C. (1984). Language curriculum development. *RELC Journal*, *15*(1), 1-14.

Richards, J. C. & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics*. Third ed. London: Longman.

Seliger, H. W. & Shohamy, E. (1989). *Second language research methods*. Oxford: OUP.

Slimani, A. (1996). Evaluation of classroom interaction. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 197-220). Cambridge: CUP.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O. & Provus, M. M. (1971). *Educational evaluation and decision-making*. Itasca, Illinois: Peacock.

Tyler, R. (1949). *Basic principles of curriculum and instruction*. New York: Harcourt Brace.

Mohammad Zohrabi
Department of English Language,
Faculty of Farsi Literature,
The University of Tabriz,
Tabriz, Iran
Phone: 00984113392163
E-mail: mohammadzohrabi@gmail.com

Received: July 30, 2012
Revised: June 19, 2012
Accepted: July 5, 2012