TEFL Program Evaluation at Master's Level in Iran

Elham Foroozandeh
Ph.D., Shiraz University

Abdolmehdi Riazi
Associate Professor, Shiraz University

Firooz Sadighi
Professor, Shiraz University

Abstract

This paper reports part of a large-scale study designed based on Stufflebeam's CIPP Model (2002) to evaluate the TEFL curriculum in MA program implemented at nine major universities in Iran with regard to the Official Curriculum developed in 1987. Participants included 68 MA students, 34 instructors, and nine administrators. Required data were collected through three questionnaires, interviews, and written responses. Two course-based questionnaires including 3-point and 5-point Likert type items as well as one open-ended question for students and instructors were developed based on the Official Curriculum. To develop the Administrators' questionnaire, several program evaluation questionnaires for administrators were consulted out of which items relevant to the Iranian educational context were selected. This questionnaire also included 5-point Likert type items as well as two open-ended questions. All the three questionnaires shared 23 items on the program's work plan. Interviews were also conducted with 18 instructors and 30 students. The data were analyzed using both quantitative and qualitative procedures, but due to the large volume of analyses, this paper reports only the results of the qualitative analysis in detail. To validate and interpret the findings, the same were discussed with about 10 MA

students and 2 instructors. The findings generally revealed that (1) there was no consensus among the participants regarding the overall aim of the program, (2) the implemented curriculum is partially compatible with the Official Curriculum, and (3) the participants generally felt the need for (a) the official curriculum's revision, (b) reform in program delivery, and (c) reconsidering the screening system.
**Key words:** Program evaluation, (official/implemented) curriculum, CIPP Model, QSR Analysis

1. Introduction

Program evaluation, from a general point of view, refers to "a structured process that creates and synthesizes information intended to reduce the level of uncertainty for stakeholders about a given program or policy" (McDavid & Hawthorn, 2006, p. 3). It involves careful collection of information about a program or some of its aspects to make necessary decisions (McNamara, 2002). Evaluation is a widely used means for assessing a program's efficiency in public, nonprofit, and private sector organizations around the world. The application of the concept dates back to the 1960s in the United States during the period of the Great Society social programs associated with the Kennedy and Johnson administrations. The *Great Society* refers to a set of domestic programs proposed or enacted in the US by President L.B. Johnson. The reform program sought two main goals, namely, the elimination of poverty and racial injustice (Rossi, Lipsey, and Freeman, 2004). Having conducted an overall evaluation, some of the programs were eliminated or had their funding reduced, while many of them, including *Medicare*, *Medicaid*, and federal education funding, continue to the present. In effect, program evaluation is so essential a process that some key organizations in the United States and Canada have developed theoretical frameworks to implement sound program evaluations (GAO, 1998). There are different approaches to evaluation and the purposes that it would serve. Stufflebeam's

CIPP (Context, Input, Process, Product) Model developed first in 1966 is a basic, most frequently used model in both education and HRD (Human Resource Development) settings. Stufflebeam's improved model released in 2002 in the form of a checklist is reproduced in Figure 1 as follows:

Figure 1: Stufflebeam's CIPP Model
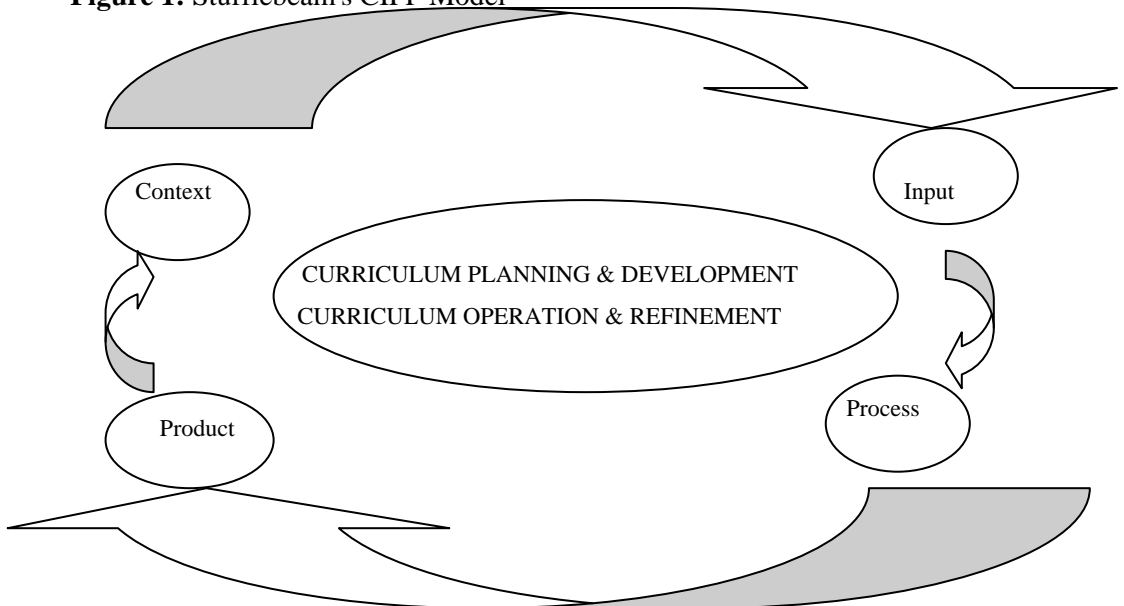
Descriptors (Stufflebeam, 2002, pp. 1-10):
- *Context evaluation*: What needs to be done? It assesses needs, assets, and problems within a defined environment;
- *Input evaluation*: How should it be done? It assesses competing strategies and the work plans and budgets of the selected approach;
- *Process evaluation*s: Is it being done? They monitor, document, and assess program activities;
- *Product evaluation:* Did it succeed? In the 2002 checklist, the Product Evaluation part is divided into **impact, effectiveness, sustainability,** and **transportability** evaluations as follows:
(a) impact evaluation assesses a program's reach to the target audience;

*(b) effectiveness evaluation* assesses the quality and significance of outcomes;

(c) sustainability evaluation assesses the extent to which a program's contributions are successfully institutionalized and continued over time; and

(d) transportability evaluation assesses the extent to which a program could successfully be adapted and applied elsewhere.

Among the earliest evaluation schema for making judgments about a program is Stake's *Final Evaluation Report* (1969 cited in Shepard, 1977) and his 1976 Checklist for *Negotiating an Agreement to Evaluate an Educational Program* based on the Final Report. Other frameworks include Shepard's (1977) *Checklist for Evaluating Assessment*, which was developed on the basis of Stufflebeam's Metaevaluation Criteria (1974 cited in Shepard, 1977), Scriven's Checklist for Evaluating Products, Procedures, and Proposals (1974, cited in Shepard, 1977), and Stake's Final Report. Another contribution was made by Heining-Boynton's (1990) *Foreign Language in the Elementary School Program Evaluation Inventory (FLES FPEI)* developed in 5 forms for students, FLES teachers, mainstream classroom teachers, students, parents, and administrators to evaluate the FLES program in the US and find whether it was improving after its decline in the 1950s and 60s.

ESL program evaluation specifically found its right footing in the field through a good number of studies in the 1980s (Jacobson, 1982; Long, 1984; Beretta & Davies, 1985; Kennedy, 1988; Brown, 1989 to name a few), and nourished prolifically in the 1990s (Lynch, 1990, 1992; Alderson & Scott, 1992; Mackay, 1994; Liskin-Gasparro, 1995; Mackay, Wellesley, & Bazergan, 1995; Kieley, 1998; Roberts, 1998; Rea-Dickins & Germaine, 1998; Anderson, 1998; and many others). Each of these studies has contributed to evaluation research in terms of defining the concept, developing a conceptual framework or model, or presenting an empirical research. For instance, Rea-Dickins and Germaine (1992, p. 10) identify two main motivations for evaluation as follows:

Figure 2: Rea-Dickins & Germaine's schematic representation of motivations for evaluation
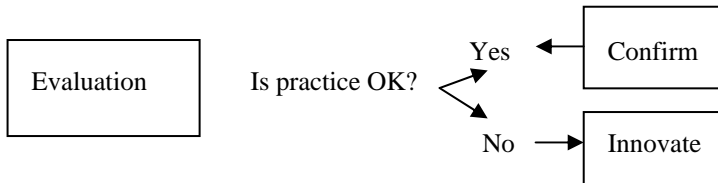
Based on this model, a program will go on as it is if evaluation *confirms* its efficiency in achieving its objectives; otherwise, there will be need for *innovation* in case shortcomings are observed.

From another perspective, Weir and Roberts (1994) distinguish between evaluation for purposes of *accountability* (at contractual and professional levels) and evaluation for purposes of program or project *development*. In this framework, accountability refers to the competency of the staff to justify the quality of their work to others, while *development-oriented* evaluation is to improve the program or the project. In educational contexts, program evaluation is concerned with both professional accountability and program development. Ingvarson's comparative study (2001) of two sets of teaching evaluation standards established by the UK and US Education Departments questions the validity of nation-wide evaluation standards mostly observed in the UK. This is what Widdowson (1990) has objected to as a global evaluation model expected to be generalizable to different educational settings. Another evaluative study carried out on teacher professional development in former Yugoslavia in late 1980s (Radulovich & Rajovich, n.d.) revealed that the pre- and in-service teacher education in the country had no structure and pattern to involve teachers from diverse backgrounds. They specifically questioned the dated *curricula* and methods of teacher education as one of the major drawbacks of the system (italics added).

Curriculum as also highlighted in the Yugoslavian study is basically a major factor in the evaluation of an educational program. According to Howell and Nolet (2000), the term could refer to a variety of variables, including the courses taught in a school or program, the documented list of courses to be taught, a set of

teaching materials organized in some sequence, or a framework for selecting and organizing learning experiences. On the basis of the wide scope of curriculum and its pivotal role to all instructional activities, they define three dimensions for curriculum, namely, the *intended curriculum* (believed to be implemented in the given educational context), the *taught curriculum* (what is actually implemented in the classroom), and the *learned curriculum* (what students actually learn as a result of being in the classroom)(p. 6).

Curriculum could also refer to the whole content to be taught and to be realized within one school or educational system. It includes methods, approaches, evaluation measure, teaching materials and equipment, "and even teacher education" (Medgyes & Nikolov, 2002, p. 196). In 1992, two very similar evaluation frameworks for educational programs were introduced which are compared in Table 1 below.

**Table 1:** Rea-Dickins & Germaine's (1992) and Alderson & Beretta's (1992) frameworks compared

Items	Rea-Dickins and Germaine (1992)	Alderson and Beretta (1992)
1	Who / why	Stakeholders / purpose
2	What	Content
3	Criteria	___
4	When	Timing
5	Procedures	Method
6	Who/Where to get information from	Sources of information
7	Who to involve	Stakeholders, [organization/department] staff, those to be evaluated, those who are the sources of data
8	How to manage evaluation	Project framework
9	What to do with the information	Implementing an evaluation
10	Constraints	Factors the absence of which causes restrictions on the progress

As for timing, program evaluation may be scheduled from different perspectives. Henning (1987) classifies the different types of evaluation into *prior-to-program implementation*, *during-program delivery*, and *following-program execution*. Similarly, McDavid and Hawthorn (2006) divide it into *ex ante* evaluation (done before program implementation) and *ex post* evaluation (done after program implementation). As for purpose, which is closely related to timing, McNamara (2002) identifies three major types of evaluation, namely, goals-based, process-based, and outcomes-based with sample questions pertinent to each type.

Along the same lines, Rossi, Lipsey, and Freeman (2004) identify five dimensions of program evaluation including (a) needs assessment, which examines the nature of the problem that the program is addressing, (b) program theory, or the program's conceptual framework, (c) process analysis, which is concerned with how the program is implemented, (d) the impact of evaluation, which determines the effect of the program, and (e) cost-benefit or cost-effectiveness analysis, which assesses the efficiency of the program in terms of the costs and benefits.

One of the most popular evaluation frameworks is the Logic Model represented in Figure 3. To clarify the nature of this approach, Taylor-Powell (2005) notes that it is not a theory or reality, nor even an evaluation model. She defines it as a "framework for describing the relationships between investments, activities, and results…" (Slide 31).

Figure 3: Generic Logic Model (Taylor-Powell, 2005)
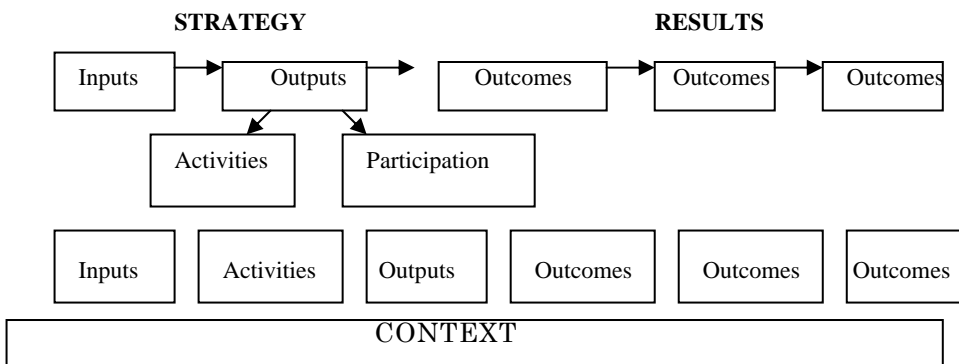
There are other frameworks that consider program evaluation (PE) in relation to performance measurement (PM), and, despite a close relationship, make distinctions between the two. Watson (2005) defines PM as the ongoing monitoring and reporting of program progress and accomplishments usually carried out by the manager(s), while PE refers to a systematic study usually carried out by internal or external evaluators to answer questions about how well a program is working to achieve its outcomes.

In addition to evaluation models, several evaluation checklists are also professionally developed to be used by evaluators for the purpose of evaluating their evaluation (Stufflebeam, 1999), to design their evaluation (Stufflebeam, 2004), to focus their evaluation on utility (Patton, 2002), or to use the checklist for qualitative evaluation (Patton, 2003).

As mentioned before, the evaluation literature includes quite a few empirical studies on Teacher Education (TE) programs (Allebone and Davies, 2000). Other descriptive studies on teacher education (Colorado State University-Pueblo, 1995; LCSC, 1999; Moore, 2000; Esche, Chang-Ross, Guha, Humphrey, Shields, Tiffany-Morales, Wechsler & Woodworth, 2005) provide full information on the goals and standards, as well as the courses and materials that are scheduled and selected on the basis of the defined objectives. Likewise, the results of Wilson, Floden, and Ferrini-Mundy's study (2002) support the effects of subject-matter knowledge, pedagogical training, and policies on the enhancement of teacher preparation. Among these studies, Darling-Hammond, Chung, and Ferlow's research (2002) is considerably useful in that it provides empirical information on the effect of Teacher Education Programs on the preparation and performance of prospective teachers.

Unfortunately, no evaluation studies have been carried out in the history of TEFL education in Iran. But in 1999, Keyvanfar evaluated the Undergraduate Translation Program in seven branches of the Islamic Azad University in Tehran and four neighboring cities. The results of her study revealed a general pattern of improvement in language proficiency and translation ability, but the

program was not successful with regard to the program's objectives, namely, (a) gaining general proficiency in all four skills of English, and (b) gaining translation skills and having adequate opportunities to actually translate different kinds of texts. Eight years later, the same study was replicated in 6 branches of the Islamic Azad University in Tehran and the same four neighboring cities (Rahmani, 2007). Findings revealed general success of the program despite a number of problems to be considered by the policy makers. Rahmani also reports that the students had either overestimated or underestimated their abilities compared with their scores on the two measures. These two studies are valuable as the first steps toward recognizing the necessity of evaluation in the Iranian higher education context.

The present study is a goals-based, process-based evaluation of the TEFL Program at Master's level in 9 major universities in Iran. It was intended to find out if the implemented curriculum at the English Departments in the participating schools is in line with the Official Curriculum and whether the stakeholders are satisfied with the program delivery and with the Official Curriculum, for that matter. Within the Stufflebeam's CIPP model (Context, Input, Process, Product), the study addressed both goals-based and process-based questions in an academic context and was of a during-program delivery type. The *product*, i.e., the program's success, was evaluated in terms of *impact*, *effectiveness*, *sustainability*, and *transportability* of the program. The following research queries emerged on the basis of the Model:

(A) Context Evaluation: What needs to be accomplished and in what context?
1. Does the Implemented Curriculum correspond to the objectives of the program specified by the policy makers in the given Graduate Department?

(B) Input Evaluation: How can the objectives of the program be accomplished? How to utilize the resources?
2. Is the budget allocated to the program sufficient?
3. Is the program's work plan and schedule sufficient and feasible?

3.1 Are the course syllabuses designed in line with the objectives of the Official Curriculum?

3.2 Do the teaching materials meet the objectives of the Official Curriculum?

3.3 What are the teaching and assessment method(s) used by the professors?

(C) Process Evaluation: What is being done and is it effective?

4. Are the course syllabuses implemented in a way to meet the objectives of the Official Curriculum?

5. What are the program events/problems that contribute to or hinder the effective implementation of the Official Curriculum?

(D) Product Evaluation: Was the program successful? (Impact, effectiveness, sustainability, transportability)

D1- Impact evaluation: the extent to which the program reaches the target audience

6. Is the program delivered to the right beneficiaries?

7. What is the stakeholders' impression of the impact of the program?

D2- Effectiveness evaluation: quality and significance of outcomes

8. Is the overall performance of the students satisfactory with regard to the instructional input and the objectives of the Official Curriculum?

9. Are the students satisfied with their achievements and the program's work plan?

10. Are the professors satisfied with the students' performance and the program's work plan?

11. Is the administrative staff satisfied with the program in terms of the objectives of the Official Curriculum, the Implemented Curriculum, and the students' performance (learned curriculum)?

D3- Sustainability evaluation: program's continuity

12. Is the program's contribution to the betterment of the academic community institutionalized and continued over time?

D4- Transportability evaluation: program's adaptability in relevant contexts

13. Is the Program successful enough to be adapted in new contexts?

2. Method

Participants

111 male and female respondents participated in this study. They included 68 MA students, 34 instructors, and 9 administrators (faculty deans, department heads, research deputies, and deputies to educational affairs) at the English Departments of nine universities in Iran (seven major universities in Tehran, Shiraz University, and Isfahan University). The students either had graduated or were in the process of writing their theses.

The participating instructors were assistant, associate, and full professors of Linguistics or Applied Linguistics with Ph.D. degrees. The administrators included department heads, deputies to educational affairs, and deputies to research affairs in the participating schools with Ph.D. degrees in Linguistics, Applied Linguistics, or other majors. Three of them were also the ex-members of the High Council of the Cultural Revolution.

Instruments: questionnaires and interviews

The instruments included questionnaires (fixed-response and open-ended questions) and interviews. Based on the Official Curriculum for TEFL Program at Master's Level approved and made effective in 1987 by the Iranian Ministry of Higher Education, three questionnaires for Instructors (99 items), Students (125 items) and Administrators (42 items) were developed. The semi-structured interview (see below) sought the respondents' attitudes toward (a) the overall aim of the program, (b) the official and the implemented curriculum and whether the latter was in line with the overall aim and with the official curriculum, (c) the instructors' teaching and assessment methods, (d) the efficiency of the screening system, the weak points, and suggestions for improvement.

Procedures:
Developing and administering the questionnaires
The Official Curriculum for TEFL Program at Master's Level, which included course objectives specified under each defined course, was first translated into English. Then, each specified objective for a given course was transformed into a question. This set of questions formed **Section I** of the questionnaire that focused on the Required Courses and Electives with three-point checking items on the Likert Scale. For **Section II**, several attitude questionnaires were consulted and items that best fit the Iranian educational system were singled out and adapted as appropriate. This section included five-point items on the Likert scale asking the respondents' attitudes toward the skills-developing and administrative aspects of the program, as well as one open-ended question asking the respondent's general impression of the program quality and how it could be improved in terms of the courses, instructors, and the method whereby the program is delivered. The students' questionnaire contained 4 more questions with sub-items in Section II regarding Supervisors and School/Faculty. Depending on the group of respondents, i.e., students or instructors, the question stems were structured in a way that they would address the right respondents. The Administrators' Questionnaire was also developed on the basis of the existing questionnaires for executives, and the relevant items were selected and modified to fit the Iranian administrative context. The reliability estimates of the three questionnaires using Cronbach's Alpha are as follows:

Administrators' questionnaire with 42 items: **r = .930**
Instructors' questionnaire with 91 items: **r = .933**
Students' questionnaire with 125:   **r = .957**

The content validity of the instructors' and students' questionnaires was checked in view of the program objectives specified in the Official Curriculum. The pilot study verified that both the course-based items and the items regarding the program's work plan, school and faculty, and supervisor, all addressed the same objectives identified at the time of planning this program

evaluation. Face validity was also improved for ease of checking and physical management of the questionnaires.

The Instructors' and Administrators' questionnaires were either given in person or emailed to the participants who expressed their willingness to take part in the study.   Some of the instructors/administrators refused to participate, as they did not believe in the usefulness of program evaluations. Others promised to return the questionnaires, but they never did. Still others objected to questionnaires as time-consuming and inefficient instruments for data collection and apologized for not participating.

The Students' questionnaires were distributed almost in the same way. However, as several of the student-respondents were in the process of writing their theses and were not available at schools, there was difficulty contacting them to explain the purpose of the study and the significance of the responses of each and every participant.   Therefore, the questionnaires along with explanation about the purpose were given to some of them in person, and emailed to others through personal correspondence or with the help of colleagues or other students who had access to the others. All the participants were assured of the high confidentiality of the responses by confirming that no personal names/information would be disclosed to any individual or organization at any stage of the research. So, the participants confidently expressed their views in the open-ended question shared among all the three questionnaires about how the program could be improved in terms of the courses (both required and electives), instructors, and the methods whereby the Program is delivered. When cases of odd checking were observed in individual questionnaires, such as checked boxes for courses which were never offered in a particular school, the respondents would be contacted and requested for clarification, and the discrepancies would be removed.

Interviews
18 instructors 3 of whom were also administrators agreed to participate in the interview. They talked about their teaching and assessment methods as well as issues in program administration, and gave general comments on the quality of the program. In case

some key points were missed in the interview for any reason, the same were emailed to the informants and the printed answers were added to the transcribed interviews. They also discussed the key issues regarding the screening system, program delivery, and program efficiency in detail. Interviews took between 30 to 120 minutes depending on the informants' available time and their willingness to share their views about the Official and the Implemented Curriculums as well as fundamental issues at the administrative level. The interviews were not taped as some instructors would prefer otherwise. So, the comments were primarily jotted down using abbreviated language at the interview sessions, and were transcribed in detail shortly after the interviews were over. Interviews with 30 students were conducted in the form of classroom discussions when possible, in conferences, and other academic events. The interviews were in English and Persian, as the participants freely switched between the two languages during the interviews. For the sake of homogeneity of the transcriptions, the interviews were all translated and transcribed in English.

3. Organization and analysis of the data

The fixed response (Likert-type) items that comprised the quantitative part of data analysis were organized and analyzed using the SPSS. The narratives including the open-ended responses and interview transcripts were organized and analyzed using a qualitative procedure described bellow.

Qualitative data analysis
The QSR analysis (QSR NVivo 7, recent software for qualitative research) was used to analyze the data collected from the narratives. The steps followed to set up the project briefly include (a) importing the narratives into NVivo 7 environment, (b) creating Free Nodes, (c) creating Tree Nodes, (d) creating Cases, and (e) creating and executing Queries.

   Free nodes included the key phrases and sentences which were directly or indirectly relevant to the research questions. To create free nodes, the key sentences or paragraphs in all the written

responses and transcriptions were boldfaced and coded, i.e., given a key label. Then, the coded ideas that focused on the same theme were grouped under a free node, e.g., "students' (dis)satisfaction with library resources." In this way, 144 free nodes (child nodes) were identified, which, based on the central theme, were grouped under the relevant *parent node* with a more inclusive label. Each parent node, therefore, would include the relevant child nodes—the free nodes previously identified. This re-coding of the ideas led to the creation of Tree nodes extended as shown in Figure 4 bellow:

Figure 4: The extended parent nodes and the child nodes screenshot from the QSR NVivo environment

Cases were created in terms of the school and students' admission year. Thus, each instructor and administrator had one case, i.e., school, and each student had 2 cases, i.e. school and

admission year. Finally, in order to obtain the most comprehensive thematic organization of ideas, *Queries* were created based on searches for key words that could best address the same issues specified in the research questions, e.g., **courses**, **program's aim**, **instructors**, etc. By *running* the queries, relevant narratives would be imported to the window specific to queries. The narratives, hence, *results*, were examined and the key words were revised so as to accommodate a larger scope of relevant ideas. To include all the relevant ideas about the same issue and have a bigger picture of each theme, query results were merged under the corresponding tree nodes.

QSR analysis provided the exact number of cases concerned with the key ideas identified in the parent nodes, and showed the frequency of each key idea cited in the narratives. For example, in 60 of the narratives, 113 cases of reference to the program's aim were identified (Figure 4 above), which is an indication of the respondents' main concern with the program's work plan. Classroom Procedures, Program Delivery, and Affective Issues are the second, third and fourth most cited issues cited by the respondents.

4. Discussion

To yield dependable results, program evaluations need to be carried out within a well-constructed framework. This study was designed based on Stufflebeam's CIPP Model (2002). The four elements of *Context, Input*, *Process*, and *Product* in the present evaluation were adjusted to the model and were defined as follows:

- **CONTEXT:** English Graduate Departments at 9 major schools in Iran. The implementation of the Official Curriculum was the main purpose.
- **INPUT:** The sources of data: stakeholders (students, instructors, administrators), the course syllabuses, and teaching materials, the screening system, whether the design and implementation of course syllabuses is in line with the objectives of the Official curriculum;
- **PROCESS:** *How* the curriculum is implemented; whether program delivery is in line with the Official Curriculum.
- **PRODUCT:**

*(a) impact***:** whether the program is delivered to the right beneficiaries;

(b) effectiveness: whether the program is effective in terms of producing the expected results (hence, students' knowledge and skills, academic interactions/exchanges);

(c) sustainability: whether the program's contribution to the betterment of the academic community is institutionalized and continued over time;

(d) transportability: whether the program is successful enough to be adapted in new contexts

Along with the adjustment of this evaluation study to the CIPP Model, research queries emerged in relation to each element. Data analysis generally revealed that:

(1) the students and instructors did not differ in their evaluation of the program's work plan which both groups rated average. Therefore, answers to the **research queries 9, 10, and 11** would be, "Students, instructors, and administrators show average satisfaction of students' performance and the program's success in developing professional and interpersonal skills, creativity, and critical thinking in students."

(2) Students and instructors significantly differed in their evaluation of the implemented curriculum. More specifically, regarding **queries 3.1, 3.2** (course design and teaching materials), **and 4** (implementation of the designed courses), significant differences were observed between the two groups' views of the program delivery in terms of the design and implementation of the 7 major courses commonly offered at all of the participating schools (Methodology, Linguistics, Testing, Seminar, Writing, Research Methods, Skills), but, school wise, no significant difference was found between the two groups. QSR analysis showed that the most frequent suggestions and objections addressed the issue of too much concern with theoretical issues and little practicality of the courses. However, examination of the Official Curriculum shows that the Program is basically theory-based, and the only practical course included in the

curriculum is the 2-credit-unit *Practicum* which is offered as *elective* in all of the participating schools despite the fact that it had been identified as "required" in the Official Curriculum. So, apart from the exclusion of *Practicum* from the required courses, the implemented curriculum at this level sounds to be in line with the Official curriculum. However, the analysis of the participants' views revealed that major revisions need to be thoughtfully made to the implemented curriculum, and the Official curriculum, for that matter. This would cast doubt on the success of the presently running program, which is adapted in almost the same manner in new contexts, e.g., the non-profit colleges newly established. Therefore, as to **query 13** (program's transportability in relevant contexts), the QSR analysis of the participants' views indicated a general concern with the need for renovations, implying that the program is not successful enough to be adapted in new contexts.

(3) Students also stressed the need for oral communication courses, as they were generally dissatisfied with their communication ability. The instructors and administrators were also concerned with the same issue, not mentioning their dissatisfaction with students' professional knowledge. But they sought the source of the problem in the screening system not in the lack of oral skills courses. Thus, the general dissatisfaction with the efficiency of the system in screening in the qualified applicants provided answer to **query 8** (overall performance of the students). That is, neither the students nor the academic staff felt satisfied with students' overall performance.

(4) The issue discussed in 3 above is closely related to **query 6**, whether the program is delivered to the right beneficiaries. In social studies, when a program is designed and implemented with the purpose of removing some ills or improving living or health conditions of a particular community, care will be taken as to deliver the program to all the stakeholders, hence, the right beneficiaries. In the case of MA Program in this study, the same is supposedly delivered to the right beneficiaries, hence MA applicants. However, as the respondents generally doubted the

qualifications of a considerable number of admitted students, questions would be raised as to whether the right beneficiaries do not have the chance of admission for one reason or another, whether the program fails to meet its objectives, whether the objectives are not appropriately defined, or whether new policies need to be made in order to, quoting one of the instructors (also Department Head,) "reinvent the wheel."

(5) Based on QSR analysis, motivation as one major drive in the stakeholders' achievements and satisfaction is noticeably decreasing. The issue was discussed in relation to the stakeholders' expectations from each other which remained non-defined or ill-defined throughout program delivery. When interviewed about the compatibility of the program with the graduates' future employment, one administrator contested that the program should not be held accountable for the success or failure of the graduates' employment. They reasoned that graduates would enter a profession not at all compatible with their major field, and this is not the concern of the MA Program. On the other hand, regarding course design and implementation, students wondered how they could improve their practical teaching skills, as most of them were either officially employed by the Ministry of Education or were employed at some language schools on contractual basis. This issue may be rooted in the discrepancy between the overall aim of the Program and the course distribution. Thus, as an answer to **query 12**, there is need for concerted efforts to be put into the betterment of the academic community, which sounds to be lacking in the Program at present. Along the same lines, as little consensus was observed among the participants regarding the program's overall aim, the student respondents generally doubted the contribution of the program to the improvement of their professional skills as language teachers. Similarly, instructors and administrators believed that the graduate students are neither trained to be applied linguists nor researchers, nor language teachers taking into consideration the courses defined and designed to be implemented in Master's

Program. So, as for **query 7** (stakeholders' impression of the impact of the program), the QSR analysis revealed general uncertainty about the impact of the program on the graduates' professional life as language teachers or applied linguists or researchers.

(6) As for **query 3.3** (instructors' teaching and assessment methods), QSR analysis of instructors' written answers and transcriptions showed that class discussions of topics specified for each session is the most frequent or favored technique in presenting new lessons.   Not all instructors had course syllabuses with defined objectives and course requirements to follow throughout the semester. Students were provided with a list of readings to be covered based on which they were supposed to take midterm or final tests. Some instructors provided their course syllabuses to contribute to this program evaluation. One interesting difference in their views of the course materials related to articles and text books. Some instructors doubted the efficiency of text books in that the same are not comprehensive enough to familiarize the students with up-to-date findings in the field. They regarded journal articles as the most effective teaching materials which present much information in few pages. Other instructors doubted the contribution of journal articles to students' background knowledge in the field arguing that without strong theoretical background, students are not likely to appreciate empirical findings presented in journal articles. 4 of the instructors suggested that specialized workshops be held across universities to both develop and increase inter-school cooperation and encourage instructors and students to reflect on their practice and improve the same.

(7) Organization of workshops and national and international conferences were the most frequent suggestions made by the respondents in line with improving the implementation of the curriculum (**query 5**). No direct reference was made to specific events hindering the effective curriculum implementation.

(8) The interviewed department heads and administrators had no

documented idea as to whether the budget allocated to the
program is sufficient (**query 2**).

(9) The **first query**, *whether the implemented curriculum
correspond to the objectives of the program specified by the
policy makers in the given graduate department* is discussed last,
as all the relevant information regarding both curriculums have
already been presented with the quantitative and qualitative
analyses of the data. The answer is "Partially." The official
curriculum is not fully implemented, because, according to some
instructors and administrators, graduate departments offer
required courses based on the specialties of the faculty members.
That is, if no specialist is available for a given required course,
the Department would replace it by an elective which will be
considered as "required" thereafter. This case is observed in all
the schools that offer MA Program in *TEFL*--a misnomer.
Therefore, evaluation of the implemented curriculum**s** against the
Official Curriculum may not yield sound and dependable results.

5.   Conclusion: SWOT Analysis (**strengths, weaknesses, opportunities, threats**)

The evaluation of the TEFL Program at Master's Level in Iran led to
a major question as to whether the diversity of the implemented
curriculum**s** in 9 major universities in the country and their partial
incompatibility with the Official Curriculum are the results of the
flaws in the latter or the incomplete implementation of the same.

This study revealed that the program objectives, course design,
and course distribution should be re-defined. The QSR analysis of
the narratives summarized in the Tree Nodes (Figure 4 above)
shows that the prior concern of the majority of the respondents was
the "overall aim of the program." Although the objective explicitly
defined in the Official Curriculum is ""to train competent
individuals to teach English in colleges and higher education
institutes, and to meet the society's needs for specialized human
resources for teaching English and research in TEFL issues"
(translated from the Persian version of the *TEFL Official
Curriculum and Program Syllabus* approved by the High Council of

Curriculum Development, Ministry of Culture and Higher Education, June 19, 1987), there was little consensus about the same among the respondents. The instructors and administrators considered the aim to be either absent or unspecified regardless of their own interpretation of the aim. Considering "Language Teachers' education and training" as the program's aim, students generally complained about little practicality of the courses, which caused their disappointment with their professional skills on the one hand, and the instructors' disappointment with students' performance, on the other hand. A natural consequence of this discrepancy is that the situation would end up in chaos, because course distribution is based on each department's *interpretation* of the program's aim, and the stakeholders' expectations of one another are not properly defined.

The SWOT Analysis implemented in this study dates back to 1960-1970, and is used to find out why a corporate planning would fail (Chapman, 1995). The analysis of strengths, weaknesses, opportunities, and threats could be applied in/to any organization, business, proposition, etc. It is useful in that it provides essential information needed for an enterprise to become aware of its weaknesses and improve its products or services based on the existing opportunities. By highlighting the threats, SWOT Analysis also helps organizations avoid further losses and/or potential causes of failure. The "strengths" and "weaknesses" are mostly related to the status quo of a program, while "opportunities" and "threats" are related to the future.

Despite the reported disagreements, the MA TEFL curriculum has a clearly defined objective—Training Competent English Teachers—which is definitely a ***strength***, a must have in any program, educational, business, health care, etc. Also, another strength and, in fact, an asset to the TEFL community in Iran is the presence of quite a few dedicated up-to-date instructors who, despite their disappointment with certain administrative issues, never hold back, and stick to high academic standards at international levels.

The *weakness,* however, is partly rooted in the incompatibility of the program objective with the courses and program implementation which was fully discussed in answering the research queries above. The findings revealed that students and instructors are generally not content with course distribution and the unclear professional development of the graduates in future. Another weakness frequently cited by students was the absence of official meetings for students' briefing into MA Program at the time of admission and lack of students-instructors quality talks which would sometimes cause serious misunderstandings.

Nevertheless, the results of this evaluation shows that if concerted efforts are made to conduct needs analysis for the program reform, both students and instructors are willing to actively take part in delineating and articulating the actual needs and shortcomings based on the graduates' future employment and professional development. Some of the instructors suggested that there be planned negotiations between the **Ministry of Science, Research, and Technology** and the **Ministry of Education** to connect school English teachers and university TEFL instructors, who presently look more like inhabitants of two remote islands! Graduate departments should *open up opportunities* for TEFL graduate students, and help them put their theoretical knowledge into practice, teach in real classrooms during their graduate studies, be observed by their Practicum instructors on regular basis, and receive due feedback to be able to face challenges as prospective teachers and/or improve their current practice.

Unless a thoughtful, research-based renovation is not funded and effected, the current imperfect program implementation and the absence of planned evaluation of any type on regular basis will be veiled *threats* to the long-term sustainability of the program. The stakeholders' concern with the present state and the future of the TEFL Program at Master's level expressed in their written responses and interviews bear witness to such potential threats. In the meantime, it is so weird to find MA students admitted to TEFL Program who have no previous teaching experience, and, when suggested to apply for teaching at a language school to enhance

their professional development, they sound reluctant reasoning that they have to concentrate on their MA courses and projects. How could the admission of MA applicants who have never experienced teaching or are not going to enter the teaching profession be justified while experienced school teachers who are admitted complain about the deterioration of their creativity and wonder why their performance had been more effective before they started their graduate studies in TEFL? Why is it that the challenges in school English programs are allegedly left unattended in TEFL Program at Master's Level? What could be a more serious threat to the program's impact, effectiveness, sustainability, and transportability when the program service is not fully delivered to the right beneficiaries? The issue seems to have been evaded for long, but cannot be avoided any more! The SWOT Analysis is presented in the following grid:

MA TEFL Program SWOT Analysis

STRENGTHS	WEAKNESSES
- Clear, articulated objective - Dedicated, up-to-date instructors - IT available almost all right at all the schools - Publications in accredited local/ international journals - Recent TEFL references accessible	- Course distribution incompatible with program objective - Few up-to-date instructors - Few instructors specialized in major courses - Inter-school differences in program implementation - Knowledgeable professors' retirement earlier than expected - Lack of inter-school constructive competition - Decrease in students' and instructors' motivation/commitment - Lack of program evaluation - Lack of instructors' time/interest/motivation for

	follow-up activities and student monitoring via electronic communication • Little students-instructors rapport and quality talks • Little practicality: too many theoretical courses boring to students • Students' dissatisfaction with lack of teaching practice, worrying about their PD
OPPORTUNITIES • Develop culture of using blackboard (electronic assignments and monitoring) • Organize specialized workshops • Negotiate with the Ministry of Education, sign contracts with schools, schedule teaching hours in real classrooms for graduate (and undergraduate) TEFL students • Revise screening system; include quality/professional interviews with applicants • Admit applicants with teaching experiences	**THREATS** • Little attention/No response to students' complaints about poor methodology, too much paper work, and little/no feedback on their term projects • Present program under threat of becoming white elephant • Loss of key staff: losing qualified instructors who either retire or leave the country • Next generation of EFL teachers and TEFL instructors not promising • Public losing trust in EFL education • Be far behind IT use in program implementation

In sum, every program must consistently be evaluated. No evaluation has been done on TEFL Program at Master's Level in Iran since the establishment of the Official Curriculum in 1987. It is hoped that this evaluation, which is the first framework-based study of this type, and therefore not flawless, would encourage the planning of evaluation projects in educational settings on regular basis to improve the structure of the present study, highlight

problem areas and pitfalls in the educational context, and encourage a well-organized endeavor to bring about constructive innovations. Ongoing evaluation both for improvement and for renovation is a must do in the TEFL program in order to prevent counter-productive outcomes and stop the invisible malignant growth of deficiencies.

References

Alderson, C.J. & Beretta, A. (1992). *Evaluating second language education*. Cambridge: CUP.

Alderson, J.C. & Scott, M. (1992). Insiders, outsiders and participatory evaluation. In J.C. Alderson & A. Beretta (Eds.). *Evaluating second language education*, 25-57. Cambridge: CUP.

Allebone, B. & Davies, D. (2000). Developing subject knowledge. In Jessel J. & Griffiths, J. (Eds.). *Study to Teach: a Guide to Studying in Teacher Education*. London UK: Routledge.

Anderson, J. (1998). Managing and evaluating change: the case of teacher appraisal. In P. Rea-Dickins and K.P. Germaine (Eds.). *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, 159-186. London: Longman.

Beretta, A. & Davies, A. (1985). Evaluation of the Bangalore Project. *ELT Journal*, 39(2), 121-127.

Brown, J.D. (1989). Language program evaluation: a synthesis of existing possibilities. In R.K. Johnson (Ed.). *The Second Language Curriculum*, 222-241.CUP. Retrieved May, 2008 from http://nflrc.hawaii.edu/evaluation.

Chapman, A. (1995). SWOT Analysis. Retrieved June, 2008 from www.businessballs.com

CIPP Model of Evaluation. Retrieved October, 2006 from http://isu.indstate.edu/gilberti/...

Colorado State University-Pueblo (1995). *Conceptual Framework: Building and Bridging Communities of Learners.* Retrieved May, 2006 from http://ceeps.colostatepueblo .edu/tep/proginfo / conceptframework.ap

Darling-Hamond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53(4)*, 286-302.

Denzin, N.K. & Lincoln, Y.S. (Eds.) (2000). *A Handbook of Qualitative Research*. (2nd Ed.). Sage Publications, Inc.

Esch, C.E., Chang-Ross, C.M., Guha, R., Humphrey, D.C., Shields, P.M., Tiffany-Morales, J.D., Wechsler, M.E., & Woodworth, K.R. (2005). *The status of the teaching profession 2005*. Santa Cruz, CA: The Center for Future of Teaching and Learning.

GAO (US General Accounting Office) (1998). *Performance Measurement and Evaluation: Definitions and Relationships*. Retrieved May, 2007 from www. gao.gov/special.pubs/ gg98026.pdf

Heining-Boynton, A.L. (1990). The development and testing of the FLES Program Evaluation Inventory. *The Modern Language Journal*, 74, IV.

Henning, G. (1987). *A guide to language testing: development, evaluation, research*. Newbury House.

Howell, K.W., & Nolet, V. (2000). *Curriculum-based Evaluation: Teaching and Decision Making*. (3rd Eds.). Wadsworth (Thomson Learning).

Ingvarson, L. (2001). Developing standards and assessments for accomplished teaching: A responsibility of the profession. In D. Middlewood, & C. Cardno (Eds.). *Managing Teacher Appraisal and Performance: A Comparative Approach,* 160-179. London UK: Routledge, Falmer.

Jacobson, P.L.H. (1982). Using evaluation to improve foreign language education. *Modern Language Journal*, 66, 284-291.

Kennedy, C. (1988). Evaluation of the management of change in ELT projects. *Applied Linguistics*, 9, 329-342.

Keyvanfar, A. (1999). The undergraduate English translation major in Iran: A program evaluation. Unpublished doctoral dissertation. Islamic Azad University, Science and Research Branch.

Kieley, R. (1998). Programme evaluation by teachers: issues of policy and practice. In P. Rea-Dickins & K.P. Germaine (Eds.). *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, 78-104. London: Longman.

Lewis-Clark State College (LCSC) Teacher Education. *To Prepare Caring Professionals Who Teach for Understanding in Communities of Learning*. (Rev. Ed.). Retrieved May, 2006 from http://www.lcsc.edu/education/

Liskin-Gasparro, J.E. (1995). Practical approaches to outcomes assessment: the undergraduate major in foreign languages and literature. *ADFL Bulletin*, 26(2), 21-27. Retrieved February, 2008 from http://nflrc.hawaii.edu/evaluation.

Long, M.H. (1984). Process and product in ESL program evaluation. *TESOL Quarterly*, 18, 409-452.

Lynch, B.K. (1990). A context-adaptive model for program evaluation. *TESOL Quarterly*, 24(1), 23-42.

Lynch, B.K. (1992). Evaluating a program inside and out. In J.C. Alderson & A. Beretta (Eds.). *Evaluating Second Language Education*, 61-99. Cambridge: CUP.

Mackay, R. (1994). Undertaking ESL/EFL program review for accountability and improvement. *ELT Journal*, 48(2), 142-149.

Mackay, R., Wellesley, S., & Bazergan, E. (1995). Participatory evaluation. *ELT Journal*, 49(4), 308-317.

McDavid, J.C. & Hawthorn, L.R.L. (2006). *Program Evaluation and Performance Measurement*. USA: SAGE Publications.

McNamara, C. (2002). *Basic guide to program evaluation*. Retrieved May, 2007 from http://www.managementhelp.org/evaluatn/fnl_eval.htm

Medgyes. P. & Nikolov, M. (2002). Curriculum development: the interface between political and professional decisions. In R.B. Kaplan (Ed.). *The Oxford Handbook of Applied Linguistics,* 195-206. Oxford: OUP.

Moor, A. (2000). *Teaching and learning*: Pedagogy, curriculum, and culture. London UK: Routledge, Falmer.

Patton, M.Q. (2002). Utilization-Focused Evaluation (U-FE) Checklist. Retrieved October 2006 from http:// www. wmich.edu

Patton, M.Q. (2003). Qualitative Evaluation Checklist. Retrieved October 2006, from http://www.wmich.edu

Radulovich, L. & Rajovich, V. (n.d.). Teacher Evaluation in Yugoslavia. Retrieved May, 2006 from http://www.centers .ex.ac.uk/ historyresource/ journal3/y.

Rahmani, N. (2007). On the efficacy of BA translation program in Iran (An evaluation study). Unpublished doctoral dissertation. Islamic Azad University, Science and Research Branch.

Rea-Dickins, P. & Germaine, K. (1992). *Evaluation*. Oxford: OUP.

Rea-Dickins, P. & Germaine, K.P. (1998). The price of everything and the value of nothing: Trends in language program evaluation. . In P. Rea-Dickins & K.P. Germaine (Eds.). *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, 3-19. London: Longman.

Roberts, C. (1998). Language and cultural issues in innovation: the European dimension. . In P. Rea-Dickins & K.P. Germaine (Eds.). *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, 51-77. London: Longman.

Rossi, P.H., Lipsey, M.W. & Freeman, H.E. (2004). Evaluation: a systematic Approach. Retrieved January 2008, from http://en.wikipedia.org/wiki/Program_evaluation

Shepard, L.A. (1977). A Checklist for Evaluating Large-Scale Assessment Programs. Retrieved October 2006, from www.wmich.edu/evalctr/checklists

Stake, R. (1976). Checklist for Negotiating an Agreement to Evaluate an Educational Program. Retrieved October 2006, from www.wmich.edu/evalctr/checklists

Stufflebeam, D.L. (1999). Program Evaluation Models Metaevaluation Checklist. Retrieved October, 2006, from www.wmich.edu/evalctr/checklists

Stufflebeam, D.L. (2002). CIPP Evaluation Model Checklist. Retrieved October 2006, from www.wmich.edu/evalctr/checklists

Stufflebeam, D.L. (2004). Evaluation Design Checklist. Retrieved October 2006, from www.wmich.edu/evalctr/checklists

Watson, Y.M. (2005). Introduction to logic modeling and performance measurement workshop. 2005 Environmental Information Symposium, National Center for Environmental Innovation (NCEI). Retrieved January 2008, from www.epa.gov

Weir, C. & Roberts, J. (1994). *Evaluation in ELT*. Blackwell.

Widdowson, H.G. (1990). *Aspects of Language Teaching*. OUP.

Wilson, S.M., Floden, R.E., & Ferrini-Mundy, J. (2002). Teacher preparation research: an insider's view from the outside. *Journal of Teacher Education, 53(3)*, 190-204.