CHAPTER TWO: REVIEW OF LITERATURE

2.1 **Introduction**

This chapter provides the literature related to this particular study. It is divided into four sections. Firstly background information on components of EFL teacher education is offered. Secondly, programmes evaluation is provided. Thirdly programmes evaluation models and approaches are discussed. Finally, previous studies on the evaluation of EFL teacher education programmes are presented.

2.2 **Components of EFL Teacher Education Programmes**

General speaking, Tamaney and Palmeri (2011) compare teacher education programme to a journey which is undertaken by both teacher-educators and teacher candidates together and which is in line with the developmental principles and programme policies.

To gain foreign language teacher competencies, foreign language teacher education programmes provide sets of courses which prepare student-teachers to one or more of these competencies. However, the division of the programme components and the content of the teacher education programmes are varied.

Despite the differences in the shape and size of the language teacher education programmes, Berry (1990) argues that all the programmes have the same five components: (1) skills component, (2) methodology component, (3) theory component, (4) subject matter component and (5) language improvement component.
The skills component involves the observation and teaching practice, in other words, the practice of language teaching; methodology component is about the philosophy of language teaching. In addition, courses on language teaching methods, language testing and materials adaptation. Methodology component is, to some extent, theoretical. The theoretical component, on the other hand, is in a stronger sense theory based and focuses on theories of language, learning and teaching. The study of language, its culture and its literature and the knowledge of teaching form the subject matter component. Lastly, language improvement component aims to advance proficiency in the target language.

According to Shulman (1987), teacher knowledge base should be varied in the programme such as content knowledge (the subject matter), general pedagogic knowledge (classroom management and strategies), curriculum knowledge (using specific programs and materials), pedagogical content knowledge (the product of the interaction of the first three), knowledge of learners and their characteristics, knowledge of educational contexts, and knowledge of educational ends (knowledge of values and purposes of teaching).

Thomas (1987) divides language teacher competence into two main components: Linguistic and pedagogic. Linguistic or language competence involves the two major divisions: system/grammar division and contextual/discourse division. System/grammar division consists of two components: formal component and conceptual component. Formal component can be further sub-divided into
phonological, syntactic and lexical well-formedness while conceptual component (i.e. whether an isolated sentence is conceptually well-formed).

Lastly, contextual/discourse division contains three components: functional appropriacy, stylistic appropriacy and informational appropriacy components.

Pedagogic competence, on the other hand, is consisted of four components. Firstly, management component includes information of skills of classroom management. Secondly, teaching component includes pedagogic skills employed in communicating language to learners. Thirdly, preparation component addresses skills involved in preparing for teaching. Lastly, assessment of component includes teacher’s ability to assess his or her performance of the various skills mentioned.

In this regard, Bear (1992) classifies foreign language teacher education programme into five elements: Language development, literature and culture, linguistics, methodology and practice teaching, and foundations of education.

American Council on the Teaching of Foreign Languages (ACTFL) (2002) categorises the components of a foreign languages education programme into six types. These components include information of language, linguistics, comparisons, awareness of cultures, literatures, cross-disciplinary concepts, knowledge of language acquisition theories and instructional practices, integration of standards into curriculum and instruction, assessment of languages and cultures and professionalism. Accordingly, the council describes a competent foreign language teacher as a teacher who 1) demonstrates language proficiency, understand linguistics and identify language comparisons, 2) demonstrates cultural understandings, understands literary
and cultural texts, and integrates other disciplines in instruction, 3) understands language acquisition, creates a supportive classroom and develops instructional practices that reflect language outcomes, and learner diversity, 4) understands and integrates standards in planning and instruction, and selects and designs instructional materials, 5) knows assessment models and uses them appropriately, reflects on assessment, reports assessment results, and 6) engages in professional development and knows the value of foreign language learning.

Richards (2008) identifies two factors affected Second Language Teacher Education (SLTE) viz internal and external. The internal factors include the need for improvement through trends and advances in language teaching while the external factors like the need of English as a global language. In addition, there are two kinds of knowledge in the field of SLTE, one of which is “knowledge about” and the other is “knowledge how”. The first one is related to teaching skills and pedagogic knowledge while the other is about knowledge of language and language learning.

Ur (1992) classifies the components of ELT education programme into two kinds. The first one is theoretical and the second one is practical. She indicates that an ELTEP should neither be purely theoretical nor purely practical. This lies in two reasons: First, the English language teacher should be an educated person and the theoretical component of the ELT programmes is necessary for professional learning. Second, the practical components are employed to assist the language student-teachers in developing their own personal theories of action. Hence, the theoretical and practical components should be integrated.
The components of ESL programmes should be different from the components of the English language teacher education programmes. That is because a native speaker’s child can use and understand most of the sounds and grammatical forms in a communicative context immediately after joining school. Therefore, an ESL teacher must know more than how to speak English, and ESLTEPs must cover studies in English linguistics, anthropology, psychology, sociology and education (Kreidler, 1987).

Day (1991) identifies four basics of ESL teacher education programmes: Content knowledge (knowledge of subject matter), pedagogic knowledge (knowledge of generic teaching strategies, beliefs and practices), pedagogic content knowledge (specialised knowledge of how to teach) and support knowledge (the knowledge of various disciplines such as SLA, sociolinguistics, research methods).

Despite the difference in the shape and size of teacher training courses in English as a foreign language around the world at both pre-service and in-service levels, Cullen (1994) classifies them into four components: the first component is methodology/pedagogical skills. In this category, the courses give practice on different methods and techniques of English language teaching and various classroom skills. Methodology, micro-teaching and practice teaching are the sub-components of this category. The second component, on the other hand, is linguistics, which includes theories of language and language learning, awareness of language systems such as phonology and the place of English in the curriculum, in the society or in the world. The third is literature component, which is more common in pre-service language
teacher education programmes. This component aims not only to increase trainee’s
knowledge of the texts but also helps them use these texts for language teaching.
Finally, the fourth component is language improvement, which aims to improve the
general proficiency of the trainees.

Pre-service English Foreign Language Teacher Education Programmes (PEFLTEPs)
aim to educate effective and competent language teachers. To achieve this aim,
English language teacher education syllabus includes the following types or categories
of teacher knowledge (Roberts, 1998):

1. Content knowledge: It refers to teachers’ knowledge of target language
   systems, their target language competence and their analytic knowledge.
2. Pedagogic content knowledge: It is the knowledge of language to be taught. It
   contains adapted content and means of communicating linguistic knowledge
   according to learners’ needs.
3. General pedagogic knowledge: It includes principles and strategies for
   classroom management, repertoire of ELT activities, the use of aids and
   resources, and formal assessment of learning.
4. Curricular knowledge: It refers to the knowledge of the official language
   curriculum (exams, textbooks, etc.) and the knowledge of resources.
5. Contextual knowledge: It refers to the awareness of the characteristics of
   learners, schools and the wider system, i.e. community.
6. Process knowledge: It comprises a set of skills and attitudes (i.e. ability relates
   to learners, peers and parents, study skills, team skills, observational skills,
classroom inquiry skills, and language analysis skills) that enables the development of the teacher.

In Yemen, the division of PEFLTEP components is not different from those put forward for ESL and PEFLTEPs in the world. A curriculum of the four-year PEFLTEPs in the colleges of education can be categorized into four components viz ELT Methodology, Literature and General Education.

2.3 Programmes Evaluation

There are many definitions of the term ‘evaluation’ in the literature. All the definitions indicate that evaluation aims at gathering information about certain aspects using many different methods to make decisions. These methods are either qualitative or quantitative. According to these two methods, evaluation can be conducted as a formative or a summative form.

Patton (2000) defines evaluation as systematic collection of information about the activities, characteristics and outcomes of programmes to make judgements about the programme’s effectiveness and/or inform decision about future programming. Payne (1994), on the other hand, compares formative evaluation with summative evaluation in relation to their purpose, audience, evaluators’ role, characteristics, measures, frequency of data collection, sample size, questions asked, and design constraints.

Programme evaluation, on the other hand, is an important method in measuring whether a programme is being functioned in reality as it is planned. It helps ensure
programme quality as it allows a programme to be improved when weaknesses are identified, and it keeps a strong and successful programme untouched by verifying its successes (Rosenbusch, 1991). Its purposes may differ in number depending on the reasons for conducting this kind of evaluation. Despite the variety of purposes, programme evaluation has two broad functions or aims. The first aim is to help improve the programme which is also called formative evaluation. The second aim is to decide whether a programme should be continued which also called summative evaluation.

The essential differences between these two kinds of evaluation, formative and summative, are defined in the purpose of the information collected and how it is used. In formative evaluation, programmes are typically evaluated during their implementation of process for progress monitoring of ongoing programmes. The aim of this kind of evaluation is to provide information about how best to revise and modify these programmes for improvement. In summative evaluation, programmes are evaluated at the end of an operating cycle to judge and decide on the effectiveness of these programmes. In this type of evaluation, findings typically are used to help decide whether a programme should be adopted, continued, or modified for improvement (Alderson and Beretta, 1992).

Besides these two kinds of evaluation, there is a third kind of evaluation, which is called illuminative suggested by Richards (2001). The purpose of this kind of evaluation is to find out how different aspects of the programme work or are being implemented. In addition, it aims to provide a deeper understanding of the processes
of teaching and learning that occur in the programme, without necessarily seeking to change the course in any way.

In this regard, Posavac and Carey (2003) define six purposes of programme evaluation. They are as follows:

1. To assess unmet needs.
2. To document implementation.
3. To measure results.
4. To compare alternative programmes.
5. To provide information.
6. To maintain and develop quality

The aim of these six purposes is to plan and improve programmes, to assess their worth and to make corrections in the on-going service.

For Kirkpatrick (1998), there three reasons for evaluating programmes. They are defined as follows:

1. To justify the existence of the training department by showing how it contributes to the organisation’s objectives and aims.
2. To decide whether to continue or discontinue training programmes.
3. To get information on how to improve future training programmes.

2.4 Programs Evaluation Models and Approaches

Many different evaluation models have emerged because using one model or approach to evaluate a programme has advantages and disadvantages. So. Although each model
has different components, they seem to cause confusion in this field. That is why various models and approaches are specialised in particular area.

There is a consensus that the use of one model or approach may not be effective. The preference is given to eclectic models (combinations of different components of various approaches). Because of experience and the results of research over the past five decades, the influence and efficiency of evaluation has become inevitable in the field of programmes evaluation.

By the 1930s, research methods focus on evaluating social programmes in a variety of areas (Rossi, Lipsey & Freeman, 2004). In other words, systematic evaluation activities are typically categorised as applied social research.

By the 1940s, more attention is paid for educational programmes on determining the extent to which a programme meets its stated objectives. During this period, the first model, the behavioural objectives, is established by Tyler (1942) in the field of curriculum development. By the 1950s and 1960s, the most widely used educational evaluation models emerged in the USA and represented in the works of Taba (1962), Wheeler (1967), Kerr (1968) and Goldstein (1993). Tyler’s model is still the basis for many other models of curriculum development in spite of the fact that these models have developed and gone beyond Tyler’s original conception of evaluation. They expand, improve and recognise newly identified needs according to Tyler’s principle of using objectives as “organisers”.

In curriculum development, models of Taba (1962), Wheeler (1967) and Kerr (1968) contributed the most to develop Tyler’s model. Taba (1962) suggests
seven steps of curriculum design and development. She tries to develop Tyler’s model to enhance thinking skills of students. Her model can be considered as a further elaboration of Tyler’s.

In the late 1960s and the early 1970s, programme evaluation earned full recognition as a specialty field in the social sciences. Because of its applications contributed not only to a science of social planning and a more rationally planned society but also to the perfection of social and psychological theories of change.

Programme evaluation had not methodological or theoretical base. Its cognate disciplines such as ethnography and psychometrics have been borrowed from the social sciences when it had been for many years as a subfield of it (Worthen, 1994). During this period, the discrepancy evaluation model is proposed by Provus (1972).

Another approach is developed by Weiss (1972) which is known today as theory-based evaluation, theory-driven evaluation, or Programme Theory Evaluation (PTE). It consists of two basic elements: an explicit theory or model of how the programme causes the intended or observed outcomes and an actual evaluation that is at least guided by the model (Rogers et al., 2000). This approach forces evaluators to go beyond treating the programme as a black box and leads them to examining why observed changes arisen from an existed programme. In addition to that other models focus on how evaluators could help educational decision makers in the evaluation process. For example, the decision-making model of programmes evaluation is developed by Stufflebeam (1973). This model consists of four types of evaluation activities: context, input, process, and product.
Context evaluation assesses the problems, needs, and opportunities presented in the educational programme’s setting. Input evaluation assesses competing strategies and the work plans and budgets. Process evaluation monitors, documents, and assesses programme activities. Product evaluation examines the impact of the programme on the target audience, the quality and significance of outcomes, and the extent to which the programme is sustainable and transferable.

Stufflebeam’s model is also reconciled with Scriven’s formative and summative evaluation by explaining that formative evaluation focuses on decision making and summative evaluation on accountability. Scriven develops the goal-free evaluation model, noted that an evaluator should not be influenced or biased by the programme developer’s goal statements and should remain as objective as possible. At the same period, Owens (1973) develops the “adversary” approach which is based on advocacy: groups of evaluators debate their opposing points of view, and try to prove their assumptions by presenting stronger cases. Stake (1976) develops the responsive evaluation model in which the information about the background conditions, ongoing events and outcomes is examined. Eisner (1977) suggests an “educational connoisseurship” concept which is also known as “art criticism” model in which evaluation is only based on qualitative data collection. Two concepts are major key factors to Eisner’s model: educational connoisseurship and educational criticism.

Educational connoisseurship involves the appreciation of the finer points of an educational programme, a talent that derives from the evaluator’s experience and background in the area. Educational criticism relies on the evaluator’s ability to
verbalise the features of the programme, so those who do not have the level of appreciation that the connoisseur can fully understand the programme’s features. The second approach is called ethnographic evaluation, whose proponents believe that it can yield a more meaningful picture of an educational programme than would be possible using traditional scientific methods (Guba, 1978).

Another popular approach is established by Patton (1978) which is called utilisation-focused evaluation. This approach deals with the concern often ignored in the evaluation findings by decision makers. It probes evaluation programme sponsors to understand why this happens and how the situation could be improved.

In the 1980s, a number of evaluation researchers advocate a new form of process: formative, naturalistic (Lincoln and Guba, 1985), ethnographic (Fetterman and Pitman, 1986) and qualitative (LeCompte and Goetz, 1982) evaluation. Cronbach (1980) focuses on the political context of decision-making process. He advocates that the evaluator should be a teacher, educating the client group throughout the evaluation process by helping them refine their evaluation questions and determine what technical and political actions are best for them. During this educative process, the evaluator is constantly giving feedback to the clients, and the final evaluation report is only one more vehicle for communicating with them. He does not believe that the evaluator should determine the worthiness of a programme nor provide recommended courses of action. At the same time these researchers develop approaches focused on how evaluation results are used, others concentrated on developing methods that place few, if any, constraints on the evaluator. Known as naturalistic or qualitative, these
approaches give the evaluator freedom to choose the methods used to collect, analyse, and interpret data. Bellon and Handler (1982) design an evaluation model aims to improve educational programmes. They argue that their model can make a unique contribution to those interested in developing and improving educational programmes.

In the 1990s numerous evaluation models emerged to evaluate programmes. In the early 1990s, Nunan (1992) proposes a framework for programme evaluation includes eight key questions to be answered:

1. What is the purpose of the evaluation?
2. Who is the audience for the evaluation?
3. What principles of procedure should guide the evaluation?
4. What tools, techniques, and instruments are appropriate?
5. Who should carry out the evaluation?
6. When should it be carried out?
7. What is the time frame and budget for the evaluation?
8. How should the evaluation be reported?

There two models of evaluation are presented by Hager and Butler (1996):

1. Scientific model, in which the evaluation is theory focused.
2. Judgmental model, in which evaluation focuses on the integration of theory and practice.

The scientific model (a traditional model) plays the most important role in educational evaluation, while the judgmental model is new, has more advantages, is better elaborated and provides more qualitative evaluation.
In the late 1990s Wilkes (1999) develops four general approaches to educational evaluation:

1. Student-oriented approach focuses on measurements of student performance.
2. Programme-oriented approach compares the course as a whole in terms of its overall objectives. In addition, it describes curriculum or teaching activities and brings together reasoned accounts of how a particular course dimension contributed to the whole.
3. Institution-oriented approach aims at grading the quality of teaching for comparative reasons.
4. Stakeholder-oriented approach takes into consideration the concerns and claims of those involved and effected by the educational programme (e.g. students).

In the same year, Wallace (1999) lists out three principal models of teacher education. First one is the craft model approach which symbolises the traditional apprenticeship approach in which the trainee works with the master and gains the skills necessary for the job by observing, questioning and practicing. Second one is the applied science model approach in which teachers, based on the discoveries made in human sciences such as chemistry, physics, biology, and linguistics are told what to do in pedagogic practice. Last one is the reflective practitioner model which depends on both scientific theory and practice, and encourages trainees to reflect on them. Via reflection, one can bring theoretical knowledge to practice or vice versa so that professional development takes place.
In the 2000s, Cook (2000) argues that programme theory evaluators used in qualitative methods cannot establish that the observed programme outcomes are caused by the programme itself, as causality can only be established through experimental design.

Priest (2001) describes five models or purposes of programme evaluation:

1. Needs assessments measure the gap between “what is” (the present state of affairs) and “what should be” (the target state that is aimed at).

2. Feasibility studies measure if the programme is likely to succeed or not, by searching for alternative approaches that might help the programme delivery.

3. Process evaluations measure the gap between the aims of the programme and its actual implementation. The programme is examined in order to determine if the existing programme delivery matches its design. This information is used to make the required adjustments to the programme during its delivery.

4. Outcome evaluations measure if the learning objectives were achieved and if the stakeholders are satisfied with the products.

5. Cost analysis measures the worth of a programme in comparison with other approaches. The decisions are being made whether this programme should continue or not.

The most predominant approaches available are as follows:
1. **Product-Oriented Approaches**

These approaches focus on the goals and instructional objectives of a programme. The purpose of these approaches is to determine to what extent they are attained. The most famous supporters of these approaches are Tyler, Hammond, Metfessel and Michael.

2. **Static Characteristic Approaches**

According to these approaches that evaluation is conducted by external experts with the purpose of determining the effectiveness of the programme.

3. **Process-Oriented Approaches**

They focus on evaluation procedures which, along with goals and objectives, can change curriculum and provide its improvement (Brown, 1989).

4. **Decision Facilitation Approaches**

They mean that curriculum evaluation should serve the purposes of decision-makers who are usually administrators. Examples of these approaches are Stufflebeam’s CIPP model (1971), the CSE (the Center for the Study of Evaluation at the University of California Los Angeles) model and Provis’s Discrepancy model (1971). Baker and
Herman (2003) propose an approach, which they call distributed evaluation, to deal with large-scale longitudinal evaluation of technology.

Grenfel, Kelly and Jones (2003) propose the Competence-Based Teacher Education model approach (CBTE Model). This approach has come as a response to the three approaches proposed by Wallace (1999). It advocates drawing up a list of competencies in order to specify what to do and what is being done in order to “meet the definition of teacher”. So, trainees and trainers should be aware of the competencies that one should have in order to be an effective teacher and be able to identify the assessment criteria of trainees’ performance. This approach is now being applied as a modified version of CBTE Model by Britain, Germany and some European countries in their teacher education programmes. It includes a list of competencies and subbing skills that a teacher should have. Therefore, the framework of teacher education programmes are generally based on at least four interrelated areas which are subject matter knowledge, understanding of learners and learning, conceptions of the practice and profession of teaching, and an initial repertoire in curriculum, instruction, management and assessment.

Richards (2008) suggests one more approach to teacher education: collaborative approach. Using this approach, teachers can improve themselves by collaboration among their colleagues, their students, researchers, and other people involving in teaching and learning process such as principals and parents.
Peacock (2009) designs an evaluation of a foreign-language teacher education programme model which focuses on determining the programme’s strengths and weaknesses, and how well it meets the needs of the students.

Johnson (2009), on the other hand, criticises Wallace’s applied science model which simply assumes that one can apply the knowledge of language and SLA to the classroom. He believes that the activity of teaching L2 itself is very crucial so that one can combine not only subject mater knowledge but also the content of L2 teaching, in other words, the way of teaching L2 in the classroom.

Another two approaches have been emerged in the Yemeni EFL teacher educator context. One is Al-Anbari’s (2014) two-stage of a dialogic sociocultural approach to academic writing for teaching writing for pre-service EFL teacher context. Al-Anbari’s approach is based on three dialogic criteria: intersubjectivity, situatedness and responsivety. The second is Ba-Udhan’s (2014) three-phase of listening comprehension approach which aims to raise the student-teachers’ awareness of listening comprehension problems, their causes and the listening strategies. These models can be classified into two major types:

1. **Objective-Based Evaluation**

It identifies objectives or goals and seeks to determine if those objectives or goals are attained. It seeks the extent to which programmes meet predetermined goals or objectives. The models related to this type are discussed below:
a. Tyler’s Behavioural Objectives Model

This model is first introduced by Tyler (1942). In designing this model, Tyler depends upon the students’ emotions, feelings, intellect, and beliefs. This model consists of four parts: (1) defining objectives of the learning experience; (2) identifying learning activities for meeting the defined objectives; (3) organising learning activities for attaining the defined objectives; and (4) evaluating the learning experiences (Tyler, 1949).

In Tyler’s model, the focus is only on the degree to which teaching and the objectives of a programme, product, or process serve as a basis for devising elements and reflecting on the degree of achievement of the objectives. The major question addressed in this kind of evaluation: “Is the programme, product, or process achieving its objectives?”

Tyler’s model aims to develop students’ behaviour as the target goal of teaching. Besides, it compares intended outcomes with actual outcomes. This evaluation model relies mainly on summative evaluation in measuring students’ achievements and ignores process; the emphasis on test outcomes and the product diverts attention from the “black box” of the treatment that is received. In other words, Tyler’s model is based on the objective-oriented theory. This model takes curriculum as a means of aiming toward an educational object. Therefore, this model is also called means-objective model. Even though Tyler’s straight line model is developed
explicitly, it has several criticisms as following (Huang & Yang, 2004): in case of
evaluation, it is not ideal. This model does not have a feedback mechanism to tell
people how to correct it; it seems to lack a procedure between evaluation and
organisation which is execution. Also, the objective under Tyler’s straight line model
has a behavioural orientation. Behavioural objectives have many advantages if applied
to curriculum design, but they have some limitations on execution. For example, they
are not applied to all subjects or the design of a subject’s content. Beside advantages,
some of the weaknesses must also be addressed. For instance, it is too narrow because
it judges the outcomes as successful or not based only on whether an objective was
achieved or not. Another weakness or disadvantage lies in the limitations of choices of
objectives. They are often limited to some behaviour which can be easily quantified,
but excludes some objectives that cannot be quantified. For example, some objectives
such as increasing sense of respect for others in children cannot be objectively
quantified. This means that much of what makes people moral or ethical cannot be
included in measurable objectives.

Tyler recommends that curriculum planners identify general objectives by
gathering data from three sources:

1. The learners.
2. Contemporary life outside the school.
3. Subject matter.

After identifying numerous general objectives, the planners refine them by
filtering through two screens:
1. The philosophical screen.

2. The psychological screen.

In Tyler’s Model, the general objectives that successfully pass through the two screens become what are now popularly known as instructional objectives. This shows that the planning done takes into consideration long-term view of outcome for the students. Second, it has the function of carefully managing objectives. Therefore, it is easy to observe the outcomes of attained objectives. In addition, it is easy to find the suitability of subject’s content, activity, and teaching methods based on the objective evaluation. Combined, many of the functions in further analysis of the model are useful for the integrity of the curriculum, and for forecasting final results. Finally, the steps in the Objective model’s application are precise and logical, and thus, can be easily followed.

b. Taba’s Model

This model is developed by Taba (1962) to enhance the thinking skills of students. Her model is known as “interactive model” or “Instructional Strategies Model”. It mainly focuses on the notions of multiple educational objectives and the planning of instructional strategies. It is considered the basis of the curriculum design. To Taba, objectives can be divided into four distinct categories (basic knowledge, thinking skills, attitudes and academic skills).
Taba’s model includes five mutually interactive elements of teaching and learning system: (i) Objectives, (ii) Contents, (iii) Learning experiences, (iv) Teaching strategies and (v) Evaluative measures. In this model, there are seven steps, including diagnosis of needs, draft of objectives, selection and organisation of content, selection and organisation of learning experience, decision of the target and means of evaluation. These seven added steps are somehow similar to Tyler’s model; however, it is further divided into two parts-content and learning experiences. Her model gives due consideration to external factors that may affect various components of a curriculum, including the vicinity and community of school's location, the school district's educational policies, the goals, resources, administrative strategies of the school, teachers' personal style characteristics and the nature of the student population.

c. Wheeler’s Spherical Model

Wheeler (1967) modifies Tyler’s straight line model as spherical model. His circular model consists of five key elements, including initial situation analysis, identification of aims and objectives, selection and organisation of contents, selection and organisation of learning activities and the assessment/evaluation process (Wheeler, 1967, cited in Huang & Yang, 2004).

Wheeler’s spherical model seems much more progressive than Tyler’s straight line model since it has two advantages. First, this model has a feedback mechanism which
provides students with ways to measure their progress or accuracy. It also sets the school objective as a final step as well as the first. Thus, the curricular model makes educators identify their objectives in their evaluative stage. This model is divided into many details as middle objectives. The aggregation of middle objectives comprises the final objective. Middle objectives can further lead to the near future objective which can be achieved within a short period. Eventually, based upon near-future objectives, could also lead to the concrete objective. This model clearly calls for the setting up of objectives. Even though Wheeler’s spherical objective model progresses beyond Tyler’s straight line model, this model also receives some criticisms. The objective under this model includes behavioural characteristics. Behavioural objectives have many advantages if applied to curriculum design, but they also have some limitations on execution. How can one measure a student’s increased smoothness in writing, for example? Furthermore, this model seems to lack a procedure between organizing and integrating learning experience content and evaluation. According to Huang & Yang (2004), this procedure is the execution of this integrated content.

d. Kerr’s Model

This model is developed by Kerr (1968). It consists of four elements: objective, knowledge, school learning experience and evaluation. In this model, the objective means the students’ expected behavioural changes after learning; these changes include perception, affection, and skills. The second element, knowledge, means to
choose and organise curriculum content so as to achieve school’s objectives. Moreover, a curriculum development design should first focus on the objective to be reached. In addition, the three elements needed to establish curriculum knowledge are unity, repetition, and order. In this context, unity means to establish a connection with the field of knowledge. Repetition means repeating of certain curriculum elements while order means every continuous experience that are established on prior experience. Together, these three elements become the leading principle for organising effective curriculum. The third element in Kerr’s model, learning experience, means the interactive effect between the learners and various environmental elements. It includes social opportunities from the school’s arrangement, the influence of the school community’s character, and relationships between teachers and students. Evaluation as the final element is meant to ensure the degree to which the objective is achieved. The standard of evaluation contains objective feasibility, content and method’s suitability, students’ needs and achievement, as well as the efficiency of teachers’ preparation. Many standard evaluations need to be modified for collecting information. In addition to objective examinations and paper commentary for evaluation, Kerr includes attitudinal scale, interview, aptitude test, multiple evaluations, investigated skills and group observations as ways to measure progress.
e. Kirkpatrick’s Model

This model is the most influential model included under the goal-based evaluation approach is proposed by Kirkpatrick (1959), (Carneval & Schulz, 1990; Dixon, 1996; Gordon, 1991; Philips, 1991, 1997).

This model is most widely used and popular for evaluating training and development programmes. It focuses on measuring four levels of training outcomes: reactions, learning, behaviour, and results that should result from a highly effective training programme (Kirkpatrick, 1994).

Kirkpatrick (1998) identifies four levels of outcomes of training which are hierarchically ordered: reaction, learning, behaviour and results. This model is especially suited the training contexts where specific outputs are of interest right from the outset. The main strength of the Kirkpatrick evaluation approach is the focus on behavioural outcomes of the learners involved in the training. The four levels of the model are:

The first level, *reaction*, programme evaluation includes two general approaches; formative evaluation which is also known as internal evaluation and summative evaluation which is also known as external evaluation. Reaction can also be summative in nature. In such cases, the aim of reaction evaluation is to determine the value, effectiveness or efficiency of a training programme and to make decisions concerning programme continuation, termination, expansion, modification or adoption (Worthen & Sanders, 1987). Summative evaluation provides programme decision
makers and potentials customers with judgements about a programme’s worth or merit (Worthen, Sanders & Fitzpatrick, 1997).

The main purpose of reaction evaluation is to enhance the quality of training programmes, which in turn leads to improve performance by measuring the participant’s reactions to training programme. This should be measured immediately after the programme. Level one evaluation should not just include reactions towards the overall programme (e.g. Do you like the programme); it should also include measurement of participant’s reactions or attitudes towards specific components of the programme such as the topics, contents, methodology, instructor etc.

The second level, learning, the purpose of evaluation is to differentiate between what they knew prior to training and what they actually learned during the training programme. In other words, learning evaluation is the measurement of the increase in the knowledge or intellectual capability from “before” to “after” the learning experience. Learning outcome can include changes in knowledge, skills or attitudes.

The third level, behaviour, aims to measure the change that occurs in the student’s behaviour/job performance due to the training programme. This performance testing is to indicate the student’s skill to apply what he/she learns in the classroom. It involves testing the participants’ capabilities to perform learned skills while on the job, rather than in the classroom. Change in the job behaviour is difficult to measure because people change in different methods at different times and also it is difficult to quantify and interpret than reaction and learning evaluation. Observation and
interview over time are required to assess change, relevance of change and sustainability of change in the behaviour of participants.

The fourth level, result, includes any outcome that most people would agree is “good for the business”. Outcomes are either changes in financial outcomes (such as positive ROI or increased profits) or changes in the variables that should have a reliability direct effect on financial outcome at the same point of the future. It includes other major results that contribute to the well functioning of an organisation.

The intention at this level is to assess the costs vs. benefits of training programme, i.e. organisational impact in terms of reduced costs, improved quality of work, higher productivity, reduction in turnover, improved human relation, increased sales, fewer grievances, lower absenteeism, higher work morale, fewer accidents, greater job satisfaction, etc. Collecting, organising and analysing level four information can be difficult, time consuming and more costly than the other three levels, but the results are often quite worthwhile when viewing in the full context of its value to the organisation.

Today, many evaluators point out that while the Kirkpatrick’s model is useful to evaluate a) whether learners like their instruction, b) whether they learn something from it, and c) whether it has some positive effects for the organisation, its weakness is that it cannot be used to determine the cost-benefit ratio of training (ROI). These modern evaluators consequently recommended adding the so-called fifth level to Kirkpatrick’s model, at least for some programmes.
f. Phillips Return On Investment (ROI)

Phillips (1994) adds a fifth level to Kirkpatrick’s model. Training in this model moves from satisfying trainees to improving organisational performance. Training/fellowships are carried out to have a positive impact on the organisation.

g. Hamblin’s Five Levels Model

Hamblin (1974) is one of the first writers to modify Kirkpatrick’s model. The first three levels in his model correspond closely to Kirkpatrick’s model. However, the final level is divided into two: organisation and ultimate value. The five levels of his model are therefore:

- Level 1: Reactions.
- Level 2: Learning.
- Level 3: Job behavior.
- Level 4: Organisation—the effects on the organisation, from participant’s job to performance changes.
- Level 5: Ultimate value—the financial effects, both on the organisation and the economy.

In this model of evaluating training programmes, it assumes that there is a cause and effect chain, which links these five levels of training effects, such that each level
leads to the next level. However, the chain may get broken at any of the links of the following effects which may not directly be a consequence of the previous.

Hamblin states that the task of the evaluator using this model is to identify whether the chain has continued through all the links and if not then identify which link has broken the chain and hence, put forward suggestions as to how to mend it.

Hamblin also outlines the differences between his model, Kirkpatrick’s model (first suggested in 1967) and another model, Warr, Bird and Rackham’s (1970) model.

h. Guskey’s Critical Levels

Guskey (2002) also modifies Kirkpatrick’s four levels into 5: his levels may be of relevance as he has “students” and educational environments in mind:

- **Level 1:** Participant reaction.

  **Purpose:** to gauge the participants’ reactions to information and basic human needs.

  **Technique:** usually a questionnaire.

  **Key questions:** is your time well spent? Is the presenter knowledgeable?

- **Level 2:** Participant learning.

  **Purpose:** examine participants’ level of attained learning.

  **Technique:** test, simulation, personal reflection, full-scale demonstration.

  **Key question:** do participants learn what is intended?

- **Level 3:** Organisational support and learning.
**Purpose:** analyse organisational support for skills gained in staff development.

**Technique:** minutes of district meetings, questionnaires, structured interviews or unobtrusive observations.

**Key questions:** Are problems addressed quickly and efficiently? Are sufficient resources made available, including time for reflection?

- **Level 4:** Participant use of new knowledge and skills.

**Purpose:** determine whether participants are using what they learned and using it well.

**Technique:** questionnaires, structured interviews, oral or written personal reflections, examination of journals or portfolio, or direct observation.

**Key question:** are participants implementing their skills and to what degree?

- **Level 5:** Student learning outcomes.

**Purpose:** analyse the correlating student learning objectives.

**Technique:** classroom grades, tests, direct observation.

**Key question:** does student show improvement in academic, behaviour or other areas?

### i. Bellon and Handler’s Evaluation Model

This model is designed by Bellon and Handler (1982). The purpose of this model is to improve educational programmes. They argue that their model can make a unique
contribution to those who are interested in developing and improving educational programmes. This model consists of four major elements which are as follows:

1. The four focus areas (goals, organisation, operations and outcomes).
2. Status descriptions.
3. Analysis activities.
4. Cumulative improvement components.

The first area is goals. It deals with the desired outcomes and expectations of the programme. It is important to take students’ educational needs into consideration in this area. In other words, programme goals should be based on clearly identified learner needs. In brief, educational goals are statements of desired outcomes which reflect the values of those involved in the educational programme. At the same time, they are the statements of expectations which give direction to future practices.

The second focus area is organisation. It aims at reviewing and analysing a variety of factors influencing curriculum effectiveness, namely the specific programmes offered, resources available and how they are used and procedures followed to handle communication and decision-making and the structure of the programme (Bellon and Handler, 1982). It is argued that other models mostly ignore the organisational effects on programme development; however, this area is very important and it gives proper attention so that it is possible for it to play a vital role in the improvement operation.

Third area, operations, focuses on the everyday functioning of programmes and decisions associated with learning experiences. Curriculum implementation is investigated to identify how well general goals are translated into actions to achieve
specific learning objectives for daily lessons. Curriculum authorities have the responsibility to look at ways to improve the arrangement of course content and skills to be covered at all levels. A study of programme may assist teachers to improve their repertoire of effective techniques to enhance learning.

The last area, outcomes, aims at identifying the effects of a programme on participants and on the educational setting in which the programme is established.

Bellon and Handler (1982) assert that both expected and unexpected programme results should be given attention, because unexpected results might have either positive or negative effects on overall programme success.

j. Kearns and Miller’s KPMT model

This model is first developed by Kearns and Miller (1997). It comprises four stages of evaluation, which has many similarities to Phillips work, as well as its evaluation stages look very similar to Kirkpatrick’s levels. In this model, Kearns and Miller argue that clear objectives are an essential component of a training evaluation model. In addition, they argue that training can only bring added value to organisations if the business is not performing effectively or there is a market opportunity which can be exploited. To identify bottom line benefits, pre-training measurements need to be in place. Only where the training is to bring someone up to the standards of the job is not necessary. In other words, the four-stage KPMT model starts at the beginning of the training cycle by identifying the business need rather than the training need. The
emphasis is on clarifying objectives from a business perspective rather than that of the trainees. The four stages of KPMT model are as follows:

- Reaction to training and development.
- Learning.
- Transfer to the workplace/behaviour.
- Bottom line adds value and measures in relation to the base level measures taken.

This model differs from some other models in belief that return to investment can only be looked at in hard terms. Kearns and Miller state that if a business of objective cannot be cited as a basis for designing training and development, then no training and development should be offered. That means the difference is in aim to provide a sort of toolkit to help evaluators work through the process of identifying bottom-line objectives by means of questioning techniques, evaluating existing training, and using process mapping to identify the added value to organisations.

**k. Organisational Elements Model**

Kaufman and Keller (1994) expand Kirkpatrick’s model to include societal contribution as an evaluation criteria. They argue that Kirkpatrick’s model is intended for evaluating training and that manufacturing organisations in particular are increasingly being called as account for societal consequences such as pollution and
safety. They indicate that the framework needs to modify due to these organisations now seek to evaluate other types of development events.

The model also includes some additions at the other levels such as the inclusion of needs assessment and planning in the evaluation, an examination of the desired or expected results, and a review of the availability and quality of resources. They contend that evaluation at all levels should be planned and designed prior to the implementation of any intervention.

Kaufman, Keller and Watkins (1995) reclassify the criterion in their model into the following six levels:

- **Level 1**: Input is similar to Kirkpatrick’s reaction level, but expands to include the role, usefulness, appropriateness and contributions of the methods and resources used.
- **Level 2**: Process also has similarities to the reaction level, but is expanded to include an analysis of whether the intervention is implemented properly in terms of achieving its objectives.
- **Level 3**: Micro (acquisition) is similar to the learning level and examines individual as well as small-group mastery and competence.
- **Level 4**: Micro (performance) links closely to the behaviour level and examines the utilisation of skills and knowledge. The focus is on application rather than transfer of skills and knowledge.
- **Level 5**: Macro relates to the results level and examines organisational contributions and payoffs.
• Level 6: Mega is an additional level which looks at societal outcomes.

There are some other training approaches and models. As it is discussed earlier that training evaluation itself is a less touched part of training and development, these methods have theoretical side but less in practical application. So these models are not discussed in detail. These are: Training Validation System (TVS) approach (Fitz-Enz, 1994), Input, Process, Output/Outcome (IPO) Model (Bushnell, 1990), Kaufman’s five level evaluation model, Mahapatra and Lai (2005) and Sadri and Synder (1995).

2. Process-Based Evaluation

The purpose of this type is to identify the benefits that result from the intervention. This type does not seek to determine if a pre-determined outcome or goal. This type’s models may be more useful in terms of thinking about the overall context and situation but they may not provide sufficient granularity. Its models may not represent the dynamic interactions between the design and the evaluation of training.

Few of these models provide detailed descriptions of the processes involved in each steps. Furthermore, these models do not address the collaborative process of evaluation, that is, the different roles and responsibilities that people may play during an evaluation process.
a. Goal-Free Model

In Scriven’s (1967) goal-free model, the evaluator does not pay any attention to the goals stated by the programmes developer and examines what is actually happening. In this model, the evaluators should not only limit themselves to the study of expected goals of the programme but also consider the possibility that there are unexpected outcomes which should be recognised and studied. This model focuses on the importance of evaluating not only if the goals meet but also if the goals themselves are worthy.

In this model, the main focus is on the actual outcomes rather than the intended outcomes of a programme. Thus, the evaluator has minimal contact with the programme managers and staff and is unaware of the programme’s stated aims and objectives. The major question guiding this kind of evaluation is: “What are all the effects of the programme, including any side effects?”

b. Stufflebeam’s CIPP (Context, Input, Process, Product) Model

This model is developed by Stufflebeam (1971). This model consists of the four phases to evaluation: Context (studies the environment of the programme), Input (provides information and determines how to make use of resources in meeting programme goals), Process (addresses curriculum implementation decisions which control the programme and is used to determine whether there is a difference between
the actual and the planned activities), and Product (determines whether the final curriculum product is achieving the goals). It incorporates formative evaluation. In this model, four types of evaluation are conducted to make decisions:

1. Context evaluation, which contributes to the formulation of goals and objectives.
2. Input evaluation, which furnishes information for making decisions on designs.
3. Process evaluation, which serves decision making on procedures, strategies or operations.
4. Product evaluation, which affords information for decision making on the termination, modification or continuation of the present programme.

In general, these four parts of an evaluation respectively are asked. What needs to be done? How should it be done? Is it being done? Does it succeed?

This model is based upon the view that the most important purpose of evaluation is to improve the functioning of a programme. It considers evaluation to be an ongoing process. In this model, information is provided to the management for decision making purposes. There is a three-step process: “delineating the information necessary for collection, obtaining the information, and providing the information to interested parties” (Ornstein & Hunkins, 1998).
c. Stake’s Responsive Evaluation Model

This model is proposed by Stake (1967). This model consists of ten steps for evaluating a curriculum (Stake, cited in Ornstein & Hunkins, 1998). The ten steps are as follows:

1. Negotiate a framework for evaluation with sponsors.
2. Elicit topics, issues, and/or questions of concern from the sponsors.
3. Formulate questions for guiding the evaluation.
4. Identify the scope and activities of the curriculum; identify the needs of clients and personnel.
5. Observe, interview, prepare logs and case studies and so on.
6. Pare down the information; identify the major issues or questions.
7. Present initial findings in a tentative report.
8. Analyse reactions and investigate predominant concerns more carefully.
9. Look for conflicting evidence that would invalidate findings, as well as collaborative evidence that would support findings.

d. CIRO Model

This CIRO model is proposed by Warr, Bird & Rackham (1970). This model consists of four aspects of training: context, input, reaction and outcomes. According to
Tennant, Boonkrong and Roberts (2002), the CIRO model focuses on measurement carried out before and after the training. The main strength of the CIRO model is that the objectives (context) and the training equipment (input) are considered. Context Evaluation focuses on factors such as the correct identification of training needs and the setting of objectives in relation to the organisation’s culture and climate. Input evaluation is related to the design and delivery of the training activity. Reaction evaluation looks at gaining and using information about the quality of training experience.

Outcome evaluation focuses on the achievement gained from the activity and is assessed at three levels: immediate, intermediate and ultimate evaluation. Immediate evaluation attempts to measure changes in knowledge, skills or attitude before a trainee returns to the job. For Santos and Stuart (2003), intermediate evaluation refers to the impact of training on the job performance and how learning is transferred back into the workplace. Finally, ultimate evaluation attempts to assess the impact of training on departmental or organisational performance in terms of overall results.

e. IPO (Input, Process, Output) Model

This model is developed by Bushnell (1990). It focuses more on the inputs to training. It is used by IBM and helps monitor employee’s progress by setting performance indicators at each stage. The stages are:
1. Input – such as the instructor experience, trainee qualifications, resources.

2. Process – the plan, design, development and delivery of the training.

3. Outputs – the trainees’ reactions, knowledge and skills gained and improved job performance.


f. TVS (Training Valuation System)

A Training Valuation System (TVS) is developed by Fitzenz (1994). It consists of a four-step process similar to Kirkpatrick’s framework at steps 3 and 4 but is categorised as “system-based”:

- Step 1: Situation analysis – this is similar to an in-depth training analysis. Like Kearns and Miller, Fitzenz suggests that the manager’s answers are continuously probed until some visible, tangible outcome is revealed and that the questions initially focus on the work process rather than the training.

- Step 2: Intervention– this involves diagnosing the problem and designing the training.

- Step 3: Impact – this examines the variables that impact on performance.

- Step 4: Value–this step places a monetary worth on the changed performance.
g. **Provus’s Discrepancy Evaluation Model**

This model is developed by Provus (1971). It consists of four components and five stages. The four components are determining programme standards, determining programme performance, comparing performance and standards, and determining whether a discrepancy exists between performance and standards. The five stages are: design, installation, processes, products, and cost. In this model, any discrepancy is notified to decision-makers, who in turn make decisions at each of the stages.

The decisions that the decision-maker can make are “to go to the next stage, recycle to a previous stage, start the programme over, modify performance or standards, or terminate the programme” (Ornstein & Hunkins, 1998).

h. **Brinkerhoff’s Six-Stage Model**

This model is proposed by Brinkerhoff (1987). Brinkerhoff’s “The formative evaluation of training process” model entails six stages of evaluation presented in a cycle. In this model, there is a sequence of training decisions and causal connections to follow.

i. **Nadler and Nadler’s Critical Events Model (CEM)**

This model is proposed by Nadler and Nadler (1994). It is an open model which recognises that individuals and organisations are complex. The proponents of this
model also make it clear that identifying and determining all variables when a programme is being designed is not always possible. They also state that the CEM cannot be used to make predictions and that there may be factors intervening between the start of the design process and the completion of the final design.

j. Pulley’s Responsive Evaluation Model

This model is developed by Pulley (1994). It focuses on the purpose of evaluation, the “responsive evaluation” model.

Responsive evaluation is a tool for communicating evaluation results more effectively by tailoring it to the needs of the decision-makers. Pulley argues that the objective of the evaluation should provide evidence so that key decision-makers can determine what they want to know about the programme.

The stages involved are:

- Identify the decision-makers so as to ascertain who are to use the information and what their stake in it is.
- Identify the information needs of the decision-makers; what do they need to know and how will it influence their decisions?
- Systematically collect both quantitative and qualitative data. Pulley argues that the qualitative data is normally relayed in the form of stories or anecdotes and gives life to the numbers.
- Translate the data into meaningful information.
Involve and inform decision-makers on an on-going basis.

**k. Peacock’s (2009) FLTEP Evaluation Model**

In Peacock’s (2009) an evaluation of a foreign-language teacher education programme model is the focus on determining the programme’s strengths and weaknesses, and how well it meets the needs of the students. The steps for the procedure are: (a) produce a set of questions; (b) establish appropriate sources of data for the setting; (c) choose and design data collection methods and instruments; (d) collect and analyse each set of data against the questions; (e) construct an account by relating each interpretation to the others.

**l. Al-Anbari’s Two-Stage of A Dialogic Socio-Cultural to Academic Writing Approach**

Al-Anbari’s (2014) two-stage of a dialogic socio-cultural to academic writing approach is proposed for teaching writing for pre-service EFL teacher context. It is based on three dialogic criteria: inter-subjectivity, situatedness, and responsivity. The purpose of this approach is to enable meaning/knowledge making that is culturally developmental and inter-culturally dialogic. The meaning of dialogic criteria (situatedness, inter-subjectivity and responsivity) served as the frame of reference for evaluating the dialogicality of the use of the criteria. For example, if a student-teacher
exhibits an ability to use a theorist's idea from a different field (philosophy in the case of the present study) to support an argument in his/her own field (ELE), he/she then uses the dialogic criterion situatedness effectively. This is because the student-teachers reflect on an awareness existing within and outside the field simultaneously. This awareness that result in the development of the culture of the field in terms of mind action i.e., the student is able to negotiate a new idea into the field dialogically. There is substantial evidence that whenever there is a growth in the use of the criteria dialogically, there is promised evidence in terms of the potential for cultural development and intercultural dialogue. For example, when a student shows a democratic interaction as he/she attempts presenting her/his ideas in relation to the theorists' (thus, exhibiting dialogic inter-subjectivity), there is an evidence of a broadening of this student's cultural perspective. The repeated use of the criteria dialogically that result in continuous occurrence of cultural development and intercultural dialogic meaning/knowledge making in their written academic discourse. Her approach comprises of two stages: diagnosis and intervention.

1. **The first stage: diagnosis:** The purpose of this stage is to collect information about the previous writing courses which are provided in the programme as well as the student-teachers’ dialogicality. The tools used in this stage include the questionnaire designed for evaluating student-teachers' perception of the previous writing courses taught them in their undergraduate level. In addition, the first task is designed for evaluating the student-teachers’ dialogicality.
2. **The second stage: intervention:** The purpose of this stage is to improve the criteria dialogically through creating conditions under which students' awareness and application of elements of dialogicality. This stage includes selecting texts, designing tasks, and talking to student-teachers while drafting their responses to the texts. In addition, it consists of two steps: (a) creating suitable writing tasks (three in all) and (b) analysing the responses produced in drafts by using the criteria are developed by Al-Anbari.

m. **Ba-Udhan’s Three-Phase of Listening Comprehension Approach**

Ba-Udhan’s (2014) three-phase of listening comprehension approach aims at raising awareness of listening comprehension problems, their cause and the listening strategies. His approach consists of three phases of listening: diagnosis, intervention and evaluation.

1. **The diagnosis phase:** The aim of this phase is to evaluate the current level of listening comprehension of the student-teachers. In this phase, three tools are used to collect data: a pre-test, questionnaire and interviews. The pre-test is used to assess the weaknesses and strengths of student-teachers’ listening comprehension dividing them into high and low level groups. The questionnaire and the interviews are used to identify the listening comprehension problems that student-teachers encountered, the causes of these
problems and the listening comprehension strategies they used to overcome these problems.

2. **The Intervention phase:** Ba-Udhan adopts this phase from Goh 2000 and organises it into 4 components. They are:

1. Discussing listening comprehension problems and useful strategies with student-teachers in the classroom.
2. Encouraging student-teachers to think aloud once they finish a listening task.
3. Providing opportunities for individual reflection.
4. Including tasks for raising student-teachers’ metacognitive awareness of listening to English texts.

3. **Immediate retrospective verbalization:** is carried out in two ways: introspection and retrospection. In the introspective verbalisation the text is paused several times and the subjects are asked to reflect on their listening process while listening after each pause. In the retrospective verbalisation, on the other hand, the subjects are asked to reflect about their listening process immediately after they have completed the listening task.

2.5 **Research Studies on Evaluation PEFLTEPs**

In this section, the review of the literature on evaluation studies done in PEFLTEPs is organised into two categories: evaluation studies on a whole PEFLTEP and others on only one particular part of a curriculum of the PEFLTEP. Both the categories discuss theoretical sides related to evaluation methods used in the previous studies addressed in this section. The studies reviewed in this section are listed out in geographical order
to exemplify the kinds of programme evaluation conducted in PEFLTEPs: in the global context, in the Arabic context and in the Yemeni context.

2.5.1 Evaluation Studies on a Whole PEFLTEP

In this section, a review of evaluation studies conducted on a thorough curriculum of PEFLTEP are presented.

To start with one of these studies is conducted by Sovann and Chomdokmai (2012) in Cambodia. The purpose of their study is to evaluate effectiveness of PEFLTEP at the National Institute of Education (NIE) and propose appropriate guidelines to improve the ELTEP. This study is based on Kirkpatrick’s (2006) four-level model of training evaluation. Quantitative data are collected through an English standardised test, a Teaching Knowledge Test (TKT) and a self-administered questionnaire from 89 student-teachers. Qualitative data are collected from five teacher-educators’ and two administrators’ semi-structured interview. Findings reveal that the ELTEP is effective to some extent; their knowledge and skill are at “Average” level, while their attitude is changed positively. In addition, their learning level is slightly improved and their behaviour is also positively changed to some extent.

According to the results of study, four dimensions are proposed as the guidelines for the NIE’s management team to consider when planning the ELTEP in the following years. These four dimensions are:

(1) Improving Learning Environment,

(2) Upgrading Training Resources,
(3) Improving Training Curriculum and

(4) Enhancing Programme Management.

Another evaluative study is carried out by Tunc (2010) aims at evaluating the effectiveness of Ankara University Preparatory School programme through the perspectives of teacher-educators and student-teachers. To this end, the CIPP (context, input, process, and product) evaluation model developed by Stufflebeam (1971) is used. A sample consists of 406 student-teachers and 12 teacher-educators attended the preparatory school programme in the 2008-2009 academic year. Data are gathered from a self-reported student-teacher’s questionnaire and the teacher-educators’ interview. Besides, written documents are examined. Trace test is used to investigate whether the significant differences among dependent variables across independent variables existed. Results of the study showed that the programme at Ankara University Preparatory School partially served for its purpose. The findings reveal that some improvements in the physical conditions, content, materials and assessment dimensions of the programme are required to make the programme more effective.

In a recent study done by Demir (2015) about the beliefs of EFL student-teachers and their teacher-educators on the effectiveness of the PELTEP at a state university in Turkey shows the ineffectiveness of the programme in preparing student-teachers to function in the EFL context and balancing teacher-centered and student-centered learning. Using Peacock’s (2009) evaluation model, data are collected through a questionnaire and semi-structured interviews administered to student-teachers and teacher-educators, respectively. The results of the study reveal that the
most problematic areas in the programme are its inability to meet student-teachers’ needs and its failure to give adequate training in English. In addition, the programme is outdated and needs updating by increasing the number of basic skills courses and addressing the issues regarding the practice. Besides, some suggestions are made to improve the PELTEP.

Another evaluative study is conducted by Coskun & Daloglu (2010). It aims at evaluating the PEFLTEP components using Peacock’s (2009) evaluation model at a Turkish university context. The study is based on data collected from teacher-educators and fourth year student-teachers through questionnaires and interviews. Findings reveal that although teacher-educators and student-teachers have similar views about some programme components, they hold different ideas about the balance among linguistic and pedagogic competences in the programme. For teacher-educators believe that the programme is not sufficient to improve student-teachers’ linguistic competence they think that the pedagogic side of the programme needs to be improved.

One more study is done in an EFL Turkish context by Cosgun-Ogeyik (2009) aims at evaluating English Language Teaching Education Curriculum of the (ELT) departments in the faculties of education in a Turkish university. To collect data, a questionnaire is administered to the last year student-teachers’ perspectives attending ELT department. The study shows that the programme implementation is considered to be consistent in their expectations in terms of outlining the aims of the teaching profession, the social objectives, and benefits obtained from the programme. Although
the programme is generally evaluated as sufficient, culture-specific courses are a lacking point in the programme.

In Hong Kong, Peacock (2009) conducts a study which aims to adopt a new procedure for the evaluation of EFL teacher training programmes based on principles of programme evaluation and foreign-language-teacher (FLT) education. The procedure focuses on programme strengths and weaknesses and how far the programme meets the needs of student-teachers. Data are collected from student-teachers, teacher-educators, and alumni through interviews, questionnaires, essays and analysis of programme materials. The results of the study show that the programme has much strength, including the teaching of pedagogic skills and promoting reflection and self-evaluation, but also has shortcomings. It should increase the amount of practice teaching and increase input in certain areas such as knowledge of teaching within the local socio-cultural context and classroom management. The study suggests that the procedure may be useful in other contexts.

In South Korea, an evaluative study is conducted by Nam (2005) that investigates the development of communication-based English language instruction in a Korean university context by (a) evaluating/critiquing a specific college English programme at Pusan National University (PNU), (b) describing and exploring perceptions regarding English instruction at the collegiate level in Korea among Korean college students and their teachers in the EFL programme, and (c) identifying the strengths and weaknesses of the college EFL curriculum of the university with respect to how closely it aligns with student and teacher perceptions of needs related
to English language. The study also explores the needs for future college EFL curriculum design and instructional development in the general South Korean context. Data are collected from students’ and teachers’ interviews. The results of the study reveal that, overall, while students generally seem to have somewhat of negative opinions; teachers seem to have somewhat of positive opinions about the effectiveness/quality of the new curriculum. The results of the study also show the possibility that the current communication-based EFL curriculum may not be aligned well with the students’ desires, due to several weaknesses of the curriculum itself and constrains inherent in the institutional system behind the curriculum.

A similar study in the Palestinian Arabic EFL context is done by Barzaq (2007) aims at evaluating an English Language teacher training programmes in the colleges of education, Gaza Strip Universities (STTP). To achieve this aim, a questionnaire is given to (202) student-teachers selected randomly from the three universities: Islamic, Aqsa, and Azhar. Results reveal that student-teachers in the three universities (The Islamic University, The Aqsa University, and The Azhar University) acquire (71.86%) moderately well prepared) competencies that would make them prepared to teach competently. There are statistically significant differences between the Aqsa University and the Islamic University in favour of the Islamic University. There are statistically significant differences in favour of gender. There are statistically significant differences between training in High basic school and secondary school in favour of (High basic school). There are no statistically significant
differences in all these domains and due to the number of visits by supervisors. In the light of these results, some implications and suggestions are provided.

2.5.2 Studies on Only One Particular Part of PEFLTEP Curriculum

This section deals with the studies conducted on practicum and English four skills.

Erozan’s (2005) study aims to evaluate the language improvement courses in the undergraduate curriculum of the Department of English Language Teaching (ELT) at Eastern Mediterranean University. To achieve this aim, Bellon and Handler’s (1982) curriculum evaluation model is employed. The participants of the study are six instructors teaching the language improvement courses and students enrolled in these courses. Data are collected through course evaluation questionnaires for students, interviews with students and teachers, classroom observations, and examination of relevant written documents such as course policy sheets, course materials, and assessment tools used in the courses.

The results of the study show that generally the language improvement courses are effective in terms of five aspects specified in the evaluation model employed in the study, as perceived by the students and the instructors. However, the students and the instructors suggest making some changes to the existing language improvement courses to make them more effective and better adjusted to the students’ needs and expectations.

Some important conclusions drawn and recommendations made are: Practice (i.e. practice and production) component in the language improvement courses should
be enhanced, a wider variety of authentic materials should be used in the courses, various methods and activities should be utilised in teaching-learning process, and intra-subject and inter-subject relationships (i.e. continuity and coherence) between or among the courses need to be strengthened.

In Kuwait, an evaluative study is conducted in the College of Education, Kuwait University, by Al-Nwaiem (2012) aims to evaluate the basic language skills component (BLSC) in the ELTEP through an adapted version of Bellon and Handler’s (1982) evaluation model. The BLSC comprises three courses, Writing, Reading and Conversation, which must be taken by new student-teachers in the first semester of enrolment in order to improve their language proficiency. A sample of the study consists of three lecturers teaching the BLSC courses and 55 student-teachers in their first year at the English Department. Data are collected by course evaluation questionnaires, semi-structured interviews, diaries, and relevant written documents. The findings of the study reveal that the BLSC has five major shortcomings: the first one is related to the physical environment of the college site and includes old buildings, a limited number of classrooms, shortages of learning and teaching resources and facilities, and insufficient library resources. Second, with regard to the aims and objectives of the BLSC, the findings showed some critical issues and the need for a certain degree of revision. Third one is related to the content and materials of the BLSC which reveal student-teachers’ dissatisfaction with some aspects of the content and materials of the BLSC, including boring and non-challenging topics and outdated textbooks. Fourth one is related to teaching methods which are found to be
traditionally oriented, applying a teacher-centred approach. Fifth one is related to evaluation which its findings show that students are extremely critical of the traditional assessment philosophy used by their teachers, which depends on a final examination that tests rote-learned materials. The study provides some suggestions for the improvement and development of the given programme in particular, as well as educational practice in general. In addition, the study proposed a model for evaluation which can be applied and modified depending on the specifications of any given context.

Focusing on the practical component in the Jordanian EFL context, Al-Magableh (2010) carries out a study in evaluating the English practicum at Yarmouk University from cooperative teachers’ and student-teachers’ perspectives. Data are collected through two five-point Likert scale questionnaires. Results show that there are four problems: the organisation, cooperative teachers, workshop and student-teacher. First, the problems related to the organisation are transport, lack of a manual for the training course, appointing too many students for supervisory and non-outlined teaching skills for the student-teachers. Second, the problems are related to the cooperative teachers such as impressionistic assessment, sensitivity and lack of the solutions for students’ problems. In addition, the absence of written training plans, lack of training courses for the trainers, infrequency in following students up and insufficient familiarisation of student-teachers with essential teaching skills. Third, the problems are concerned with the workshop such as focusing on familiar theoretical pedagogical issues, and lack of training in using teaching aids. Finally, there are
problems related to student-teacher like using Arabic in the classroom, teaching skills to be acquired, academic qualification and classroom management.

Another an evaluative study is carried out in an Egyptian EFL context by Elghotmy (2012) aims to investigate the microteaching practices of Egyptian pre-service English teachers in an EFLTEP probing into the perceptions of Egyptian EFL teachers and their teacher-educators of the new microteaching course. It investigated a) student-teachers and teacher-educators' perceptions about the course’s focuses; b) student-teachers and teacher-educators perceptions of the difficulties encounter during course application and finally c) student-teachers' perceptions about the helpfulness of the microteaching course in enhancing their teaching performance during practicum. Moreover, the study aims to develop a framework for microteaching curricula planning and design. Data are collected through questionnaire distributed to 10 teacher-educators and 125 EFL student-teachers in the third year English department at the Faculty of Education, Menoufia University. In addition, semi-structured in-depth interviews are conducted on 7 teacher-educators and 15 student-teachers. The findings of the study indicate that there are nineteen focuses of the microteaching course at Shebin Elkom Faculty of Education. These focuses are classified into three main themes: lesson planning skills, lesson implementation skills, and lesson evaluation skills. Results reveal that teacher-educators devote less time to practice lesson evaluation skills. Findings also reveal that EFL student-teachers and their teacher-educators encounter the following difficulties during peer group work: modeling the skills, planning a micro-lesson, teaching a micro-lesson, and when
giving and receiving feedback. Furthermore, it is found that the course is helpful in enhancing student-teachers’ teaching performance at practicum concerning their professional skills rather than their personal qualities. Psychological, socio-cultural, and socio-political factors affected the implementation of the micro-teaching course are discussed. Implications and suggestions for further research are provided.

One more study in the United Arab Emirati context focusing on the practical component is done by Ismail (2011). It investigates the views of sixty-one female student-teachers from the English Language Education Programme, Faculty of Education at United Arab Emirates University regarding the microteaching component that is offered in two courses of English language teaching methods. Both questionnaire and focus-group interview are used in collecting data. Results find that the student-teachers are very positive about the effect of micro-teaching practices on their language and on increasing their enthusiasm towards the teaching methods' courses and the focus-group interview results also confirm student-teachers’ positive attitudes towards microteaching.

In the Sultanate of Oman at College of Education, Sultan Qaboos University (SQU) by Al-Mahrooqi (2011) aims at investigating EFL student-teachers’ perceptions of competency of their training programme. To collect data, a questionnaire is used that included open-ended questions focusing on three main components: general perception of the programme and the teaching practice components (school practicum), view on classroom observations, and on feedback from their supervisors. The study shows that the student-teachers value the teaching
practice component; they have reservations about such matters as the disrespectful behaviour of student-teachers, lack of cooperation of some school teachers and of school administrations. The student-teachers also value supervisor input and feedback while some feel that there is a minimum opportunity to discuss the feedback. However, almost all the student-teachers feel adequately prepared to enter the workforce after two semesters of teaching practice.

Likewise, an evaluative study is conducted by Al-Asmari (2011) aims to evaluate the prospects of integrating technology in pre-service EFLTEP at Saudi Arabian University. For collecting data, questionnaires and interviews are administrated to (180) EFL male student-teachers, academic EFL teacher-educators and policy makers. The results of the study show that pre-service EFL teachers report average levels of technology expertise, and their general use of technology outscored their use for EFL learning purposes, academic EFL teacher-educators lack sufficient technology expertise and their technology use is more idealistic than realistic and Universities policymakers should attempt to bridge the technology gap existing between national policies and policies made at the University level.

In the Yemeni quasi-study context focusing on raising EFL student-teachers’ awareness and performance in listening comprehension is conducted by Ba-Udhan (2014).

Ba-Udhan’s quasi-study explores effectiveness of a three-phase listening comprehension approach. It includes an Awareness Raising Programme (ARP) consisted of three phases of listening: diagnosis, intervention, and evaluation. Data are
gathered through a random sample of 71 (40 male and 31 female) first-year B.Ed. EFL student-teachers from Seiyun College of Education and Seiyun College of Education for women, Hadramout University.

In the diagnosis and evaluation phase, a test and a questionnaire are used for examining whether the ARP is effective in increasing student-teachers’ awareness and in improving their performance in listening comprehension due to the increased awareness. Semi-structured interviews with student-teachers are conducted too in the diagnosis phase. In the intervention phase, classroom discussions, discovery listening method, and immediate retrospective verbalisation and classroom tasks are carried out inside classroom as well as listening diary assignments done as homework for three months. The results of the study reveal that the most common listening comprehension problems are classified under the perception phase such as difficulty in following the rest of the text, in quickly remembering the meaning of familiar words, in segmenting sounds into meaningful words or phrases, and in distinguishing between two similar-sounding words. The results of the study also show the causes behind these problems and that meta-cognitive strategies are more frequently and effectively used than cognitive and socio-affective strategies used by student-teachers to overcome their listening comprehension problems. This displays greater improvements in their listening comprehension after receiving the ARP due to their increased awareness. Finally, the study provides some implications for teacher-educators, student-teachers, and course book developers.
Another quasi-study in the Yemeni context is conducted by Al-Anbari (2014) explores effectiveness of dialogicality in student-teachers' written discourse and its effect on cultural development and intercultural dialogue. To achieve this aim, a new dialogic approach is adopted for teaching academic writing.

The purpose of her approach is to enable meaning/knowledge making that is culturally developmental and interculturally dialogic. Then, she chooses 14 student-teachers from both third and fourth year, 5 from third year and 9 from fourth year studying in department of English, Zabied College of Education Hodiedah University in Yemen. Sample is selected based on their grades (from good to excellent) in all the five writing courses done in their undergraduate level (i.e. 1,2,3,4 and Advanced Writing Skills).

For collecting data, a questionnaire, three selected summarised texts of three philosophers' views about morality, and three tasks designed are used. The study consists of two stages. In first stage, the questionnaire which is used for evaluating student-teachers' perception of the previous writing courses taught them in their undergraduate level. In addition, the first task is administered to evaluate their dialogicality. In the second stage taken about three months, an intervention is designed to create conditions under which students' awareness and application of elements of dialogicality are examined. This stage is of two steps: (a) creating suitable writing tasks (three in all), (b) analysing the responses in drafts produced using the criteria developed by her. The findings of study show that students' use of the criteria dialogically improved gradually from responding to the first task to the third task.
Moreover, it is considerable evidence of the correlation between the increase of the dialogicality of the criteria and improvement in meaning and knowledge making that is culturally developmental and inter-culturally dialogic in students' written discourse. Thus, it approves that the proposed dialogic socio-cultural approach based on the criteria is valid for teaching academic writing for meaning and knowledge making that is developmental and dialogic.

Al-Muslimi (2011), on the other hand, carries out an evaluative study on the pre-service EFL Teacher Professional Preparation Programme at Faculty of Education, Sana'a University in the Light of Quality Assurance Standards in methodology 1&2 courses, curriculum course and practicum1&2 courses. Qualitative and quantitative data are collected through three tools: a questionnaire, observation checklist and evaluation sheet, which are done with (53) fourth level English student-teachers, (65) English teachers in public schools and (6) teacher-educators in the English Department. The study shows that the level of applying most quality assurance standards in the objectives, evaluation in general, content, methods of teaching and practicum two is not satisfactory. The study also reveals that the standards of online teaching skills and technology and ethics are not at all applied.

Lastly, Al-Mohamady’s (2003) study aims to evaluate the undergraduate third–year English student-teachers’ speaking abilities in the Faculty of Education at Taiz University, Yemen. Data are gathered from (30) undergraduate third–year student-teachers through a questionnaire and the test proficiency and (6) speaking teaching staff attending the English Department at Taiz University through interviews. The
study reveals that the student-teachers are poor in performing some language functions in their conversations, because of inadequate materials that were not enough to enable the student-teachers to express themselves in English, the teacher-educators who were responsible for the selection of the learning materials are not well experienced to do the job. In other words, the topics selected for the curriculum failed to satisfy the student teachers’ need to use English as a means of communication and the teaching staffs’ methods of dealing with the relevant tasks can, thus, be shown as a decisive factor.

2.6 Relevance of the previous researches to the present study

It is clear that all the previous studies investigated several points in common. Most of them used questionnaire and semi-structure interview in order to achieve their aims. There are studies adopted new quasi-approach in improving writing skills such as (Al-Anbari, 2014) and in developing listening skills like (Ba-Udhan, 2014). They dealt with ELTEPs acceptance, training and types of roles expected of teacher-educators as well as student-teachers. Throughout the review of such studies, a number of remarks are discussed the construction, aims and methods:

- Some previous studies dealt with topics such as the changes of beliefs and perceptions; change of their linguistic competence and classroom behaviour; during teaching practice (Demir, 2015, Nam 2005, Al-Mohammady, 2003).
• Some previous studies addressed topics such as practicum, and microteaching (professional component) such as (Elghotmy, 2012, Al-Muslimi, 2011; Al-Magableh, 2010; Ogeyik, 2009; Ismail, 2011).

• Some previous studies investigated topics such as integrated technology like Al-Asmari’s study in 2011.

• Different studies are concerned to ensure the competencies of education necessary for a high quality teacher of the student teacher training programme (Al-Mohammady, 2003; AlMahrooqi, 2011; Barzaq, 2007).

• The focus in the studies was on student-teachers and teacher-educators, classroom teachers/mentors, graduates and cooperative teachers.

• The most obvious focus of previous studies in PEFLTEPs evaluation is on the teaching language skills. For example, some studies attempted to evaluate language improvement courses (e.g. Erozan, 2005). Other previous studies investigated the methods used in teaching English language skills such as some studies that compared the modern methods vs. traditional methods (e.g. Ba-Udhan 2014, Al-Anbari 2014, and Mohammady 2003). However, little is known from previous studies regarding designing a new model for evaluation of PEFLTEPs.

• This study adopts a new evaluation model that is different from previous evaluation models that are either too broad by addressing many areas of EFLTEPs such as Peacock’s (2009) model, Stufflebeam’s model and four-level model. While some of them used Stufflebeam’s model and four-level
model in evaluating PELTEPs, this study only tests this new model designed in evaluating the effectiveness of the PEFLTEPs in the colleges of education at Aden University. Last but not least, it is different from other studies which emphasised evaluating PEFLTEPs and revising from perception of student-teachers only, this study investigates the PEFLTEPs evaluation and revision from both teacher-educators’ and student-teachers’ perceptions depending upon questionnaire, interview, open ended questions and essay questions. Moreover, it analyses courses contents of the PEFLTEP.

- The studies on the career evaluation of EFL teacher education programmes focus mainly on the initial years and generally the first year of service, since the beginning years are more difficult for the teacher-educators and since the attrition rate is higher in those years. Thus, the qualitative case studies shed light on the complexities of the first year(s) of service.

- The findings from the interviews and observations of individual beginning teachers show similar results independent from where they teach. Moreover, comparing the studies conducted on beginning and experienced teachers demonstrates that the initial years of teaching are full of instability and insecurity but flexibility for change, whereas in the later years teachers become more stable, feel more secure but are less willing for trying new things.

- The present study shares several points with these studies yet it is descriptive of the PEFLTEPs at the colleges of education at Aden University aiming to develop those programmes.
Some studies emphasise knowledge, some focus on skills, some emphasise attitudes and others emphasise multiple learning outcomes. The Evaluation should focus on measuring what is covered in the programme i.e. learning objectives, content, methodology and evaluation.

2.7 **Overview of the chapter**

According to the review of the evaluation models in the literature it is concluded that:

- These models all provided evaluators with alternative ways of evaluating programmes and suggested various aspects of programmes that can be focused on evaluation whether it can be the whole programme or any aspect of the programme. Therefore, it was up to those who conducted the evaluation study to decide on the most appropriate model for their own situations and purposes.

- There is no one the best model in conducting a programme evaluation. Much depends on the purpose of the evaluation, the nature of the programme evaluated, the conditions in which training is held, the institutional and trainee needs, timescales, resources available as well as the practicality of the programme to be implemented.

- No matter which of the models is used, there is always a necessity to evaluate training programmes just as there is a necessity to evaluate any other programme.

- There are many different models of training that training programmes adopt depending on the purposes of data collection and on the types of decisions
which ultimately evolve from each purpose, the two main types of evaluation are realised in the literature: formative evaluation and summative evaluation.

It is noticed from the previous studies on the PEFLTEPs that evaluation of such programmes is important, since it identifies the weak and strong aspects of those programmes and tries to improve them through the suggestions and recommendations provided according to the results of these studies. Such conclusions are as follows:

- Research about the proper balance between content and pedagogy in the preparation of English student-teachers argue that the distinction is artificial and actually detrimental to the improvement of teaching.
- The methods used in such studies are based on cross-sectional and cross-sequential approaches. Some of them, approximately, followed the descriptive and experimental approach of research in order to achieve their purpose.
- The studies that are reviewed in this section also show this gap, despite the small number of studies found in the literature.
- The studies about current PEFLTEPs in Yemen are limited.