

# LANGUAGE PROGRAM EVALUATION: DISTINGUISHING ‘EVALUATION’ FROM ‘ASSESSMENT’

Muhammad Salim Tufail<sup>1\*</sup>, Mohamed Amin Embi<sup>2</sup>

<sup>1</sup>Language Centre, Defence University of Malaysia, Sungai Besi Camp, 57000 KUALA LUMPUR, Malaysia

<sup>2</sup>Faculty of Education, National University of Malaysia, 43600 UKM BANGI, Selangor, Malaysia

\*For correspondence; Tel. + (60) 3-90513400, E-mail: [salim@upnm.edu.my](mailto:salim@upnm.edu.my)

\*For correspondence; Tel. + (60) 3-89215555, E-mail: [m.amin@ukm.my](mailto:m.amin@ukm.my)

**ABSTRACT:** This concept article examines as well as differentiates the meaning and usage of the terms ‘evaluation’ and ‘assessment’ in the field of ELT in general and Language Program Evaluation in particular. The purpose of this article is to promote an accurate in-context understanding and usage of these two terms. It is found that while leading experts in the field have clarified these two terms as distinct from each other, they somewhat have been used ambiguously and interchangeably. Although their literal meaning portray similarities, developments and current practice in the field of ELT in general and Language Program Evaluation in particular, have led leading experts to define them as different from each other. Evaluation is accepted as the broader concept which involves judging the worth or value of an entire language program, which is made up of a plethora of elements. In contrast, assessment is specifically focused on the measurement of individual learners in terms of their language ability. While the results of assessment can be used as one of the forms of information for an evaluation, it can also be used for other purposes. It is hoped that this article will generate an awareness by providing a clearer understanding of the two key concepts, as well as serve to reconcile ambiguity, and thereby facilitating a more efficient and effective sharing of knowledge in the collective quest for further progress in the field.

**Keywords:** Evaluation, Assessment, Language Program, English Language Teaching (ELT)

## 1. INTRODUCTION

In the field of English language Teaching (ELT), the terms ‘evaluation’ and ‘assessment’ are sometimes used ambiguously [7]. Norris said that ‘one challenge for understanding the potential contribution of program evaluation within applied linguistics relates to confusion regarding the term evaluation’ (p.169-170)[11]. This situation can be perplexing for those involved in the field of English Language Teaching, particularly in the specialized area of Language Program Evaluation. The relevant experiences and reflections of the authors in the field of ELT suggests that there is a need to deliberate, clarify and distinguish between the two key concepts, particularly in looking at the differences and relationships between them. Leading experts in the field like Norris (p.576) [13] warned of the risk that might be taken if key terms like evaluation and assessment are not defined; “When we fail to define our terms, we risk talking at cross purposes, missing the point, and further alienating teachers and learners”.

This concept paper examines as well as differentiates the meaning and usage of the terms ‘evaluation’ and ‘assessment’ within the ELT sphere in general and Language Program Evaluation in particular. By doing so, this paper also contributes in adding to the relative scarcity of literature on Language Program Evaluation [12].

## 2. THE LITERAL MEANING OF ‘EVALUATION’ AND ‘ASSESSMENT’

This paper begins by examining the literal meaning of the two terms. The literal meaning of the terms ‘evaluation’ and ‘assessment’ is shown in Table 1.

The similarities in the literal meaning of the terms ‘evaluation’ and ‘assessment’ is one of the reasons why the two terms are sometimes used interchangeably. Literally, they appear to have the same essential meaning. Both involve the forming of opinions, careful thought, considerations and making judgements about some entity. However, in the field

of ELT, there is a distinct line between the two as well as overlapping areas.

**Table 1. The literal meaning of the terms ‘evaluation’ and ‘assessment’**

Evaluation	Assessment
“the act of forming an opinion of the amount, value or quality of something after thinking about it carefully.” Oxford Advanced Learners’ Dictionary [15]	“an opinion or a judgement about somebody/something that has been thought about very carefully.”
“to determine the significance, worth, or condition of usually by careful appraisal and study.”	“the action or an instance of making a judgment about something: the act of assessing something.”

Merriam-Webster [9]

## 3. THE EVOLVING CONCEPTUALIZATION OF EVALUATION: SOME HISTORICAL PERSPECTIVES

One can reflect on the way the understanding and conduct of evaluation has developed over the decades to appreciate another reason why the ambiguity and the interchangeability of the two terms ‘evaluation’ and ‘assessment’ came into place. Looking through recent decades, one may see how the definition of the term ‘evaluation’ has evolved significantly along with related developments to gradually embrace its changing conceptualizations, scope and purposes. It has since incorporated wider dimensions to encapsulate the progress made in the field of Educational Evaluation in general and Language Program Evaluation in particular.

The evolution (and variety) of definitions of the term ‘evaluation’ is, among others, a consequence of the growth

and advancement in the field over the years as experts have developed a greater variety of methods and approaches of evaluation to meet its broadening scope and purpose. Such a scenario was also tandem with the changing environment, which involved the increasing demands of stakeholders and the public, as well as developments in language education in general. This emerging trend was mentioned by Kiely and Rea-Dickins, where they said that the study and practice of evaluation has developed in diverse ways over recent decades. They went on to say that “the history of evaluation over the past several decades illustrated a progression from reliance on stripped-down statistical representations to inclusive, multi-perspective approaches” (p.6) [6].

Also mentioned by Kiely and Rea-Dickins, was that evaluation many decades ago was primarily concerned with course outcomes, particularly test results, which can also be referred to as assessment data [6]. In congruence with this, Nevo mentioned that the perceptions of educational evaluation have evolved a long way along with the changing environment since its initiation by Ralph Tyler more than half a century ago, which started with the evaluation of student achievement data [10]. In similar vein, Brown when discussing the evolution of approaches to Language Program Evaluation, began (chronologically) by describing the product oriented approach, “which focus on the goals and instructional objectives of a program with the purpose of determining whether they have been achieved” (p.224) [2]. It can be clearly seen that in the early era of evaluation, the primary concern of evaluation, which was assessment data, had led one to view these two terms (‘evaluation’ and ‘assessment’) as “similar”, “interchangeable” or even at times “referring to the same thing”.

**4. THE CHANGING SCOPE OF EVALUATION: SOME HISTORICAL PERSPECTIVES**

In line with associated developments, particularly with the desire to understand better the holistic nature language programs amid increasing stakeholder demands, the practice of evaluation has expanded considerably beyond its initial preoccupation with only assessment data or outcomes. It now encompasses the many elements that constitute to the complexity of language programs which involve both product and process dimensions and are crucial in facilitating a holistic understanding on how language programs work, particularly for the benefit of experts, practitioners and increasingly discerning stakeholders.

In congruence with this, Stufflebeam and Shinkfield mentioned the need for a much broader conceptualization of evaluation involving a combined study of performance and values [21]. Meanwhile, Weir and Roberts mentioned about embracing both the accountability and development dimensions where by doing so, it would not only measure the outcome of educational products, but also help us comprehend why things turn out the way they do in language programs [22].

The expanding scope of evaluation was also mentioned by Richards, who talked about evaluation as “an examination of the context in which the program occurs” (p. 286)[19]. We can appreciate that the context of a language program is far-reaching and comprehensive in nature and involves many program elements which affect assessment results and learner

learning. The context involves many circumstances that form the setting for a program which is made up of a variety of related elements. In this regard, Richards also said that that the focus of evaluation must be on the analysis of the different factors that determine the successful design and implementation of language programs and language teaching materials [19]. He added that “It focuses on collecting information about different aspects of a language program in order to understand how the program works and how successfully it works, enabling different kinds of decisions to be made about the program, such as whether the program responds to learners' needs, whether further teacher training is required for teachers working in the program, or whether students are learning sufficiently from it” (p. 286). On a similar note, Rea-Dickins and Germaine (p. 5) [17] said that the scope of evaluation covers the many elements which make up the entirety of a language program:

“It is important to be sure when we mention the need to evaluate our language teaching methods, our materials, our effectiveness as teachers and so on, that we actually know what it is we are evaluating. How materials are presented to learners, the types of learning tasks used, and the way we that design our courses, all for different aspects of our work as teachers. They are all part of the curriculum, of the full range of activities which take place both prior to and during the implementation of a learning programme. And they must be evaluated.”

The extensive nature of evaluation is reflected by the many aspects within the entirety of a language program which can be evaluated (with the diverse reasons for doing so is) as shown in Table 2. The congruent viewpoints from the leading experts have contributed to the more comprehensive and wider-ranging approaches of evaluations which were needed provide a deeper understanding of the myriad of dynamic and interrelated elements in a language program cause it to either succeed or fail, in order to provide informed judgements or solutions.

This shift in the practice of evaluation was much needed to meet the requirement of educational institutions and practitioners to understand the link between processes and outcomes within particular contexts, in light of the more discerning needs of stakeholders and users amidst a growing era of tighter public funds coupled with the demands for greater accountability and quality in education. The necessity to “do more with less” led to the desire to seek a greater understanding of program workings in order to make informed judgements which contribute to program improvements, became the order of the day.

**Table 2. The different aspects of a language program which can be evaluated [19. 22]**

Focal points	Why evaluate it?
Curriculum design	To provide insights about the quality of programme planning and organisation.
Syllabus and programme content	To find out how relevant and engaging it was, how easy or difficult, how successful tests and assessment procedures were.
Classroom	To provide insights about the extent to which

<b>processes</b>	the program is being implemented appropriately.
<b>Materials of instruction</b>	To provide insights about whether specific materials are aiding learner learning and teachers' teaching, and how effective is the selection of materials.
<b>The teachers</b>	To find out how they conducted their teaching, what they taught and their perceptions of the programme.
<b>Teacher training</b>	To assess whether teachers have received adequate training.
<b>The learners</b>	To find out what/how well they learned from the program, their perceptions of it, and how they participated in it.
<b>Monitoring of learner progress</b>	To conduct formative (on-going) assessments of what is being learnt by learners.
<b>Learner motivation</b>	To provide insights about the effectiveness of teachers in aiding learners to achieve goals and objectives of the program and institution.
<b>The institution</b>	To find out what administrative support was provided, what resources were used, what communication networks were employed and how effective they were.
<b>Learning environment</b>	To provide insights about the extent to which learners are provided with a conducive and responsive environment in terms of their educational needs.
<b>Staff development</b>	To provide insights about the extent to which the institutional system provides the staff opportunities to increase their effectiveness.
<b>Decision making</b>	To provide insights about how well the school staff make decisions that result in learner benefits.

In related developments, Kiely and Rea-Dickins also talked about the extensive nature of evaluation, saying that evaluation is a form of inquiry, which ranges from research to systematic approaches to decision-making” [6]. They also spelled out the many meanings and purposes of evaluation from broad-ranging perspectives from the gauging of learner achievement to program processes and outcomes, as follows:

- i. It can be part of the novice teacher’s checklist to guide the development of initial lesson plans and teaching practice.
- ii. It can be a process of determining learning achievements or learner satisfaction.
- iii. It can be a dimension of the analysis of data in a formal evaluation study.
- iv. It can refer to judgements about learners by teachers and by external assessors, external monitors and inspectors.
- v. It can be about the relationships of different program components and epistemologies developed by the people involved in programs; and the processes and outcomes which are used to show the values of a program – accountability – and enhance this value – development.

Due to its extensive and complex nature, particularly in terms of its expansive scope and purpose, the term ‘evaluation’ has taken a broad meaning, covering a wide range of the many elements which make up the intricacy of a language

curriculum. So much so that is difficult to come to a verdict on what is “the” definition of evaluation. Language Program Evaluation has thus developed far beyond mere accountability or outcome only concerns. As mentioned by Norris, it now provides a more comprehensive scope which covers the provision of “framework, heuristic, and methodologies for making sense of language teaching and learning (and other language-related endeavours) in situ, as well as for answering ‘what works’ questions and informing practical efforts at improvement” [11].

### 5. DIFFERENCE IN THE PURPOSES OF ‘EVALUATION’ AND ‘ASSESSMENT’

Linked to the broad scope of evaluation which covers the many elements of a language program, is the similarly wide-ranging nature of its purposes. This is in contrast with assessment, where it’s purposes are much narrower and are mainly focused around individual learner achievement. The purposes of evaluation encompass a much more comprehensive scope, of which assessment is likely to be one of them. The variety of purposes of evaluation is a reflection of its association with the entirety of a language programme. Leading experts talked about the purposes of evaluation from the following aspects [5,11,17,22]:

- i. Evaluation for accountability. This involves the answerability of the staff to others regarding the quality of their work. It is also carried out to determine whether there has been value for money. It is usually done to meet the demands of stakeholders in determining whether a particular programme should be carried on, discontinued or modified. This type of evaluation is largely the domain of policy makers or resource providers. Following are the main types of accountability:
  - a. Contractual accountability. This is where job descriptions and programme outcomes are clearly spelt out in formal contracts. Accountability evaluation is carried out to assess to which extent the programme outcomes have met the objectives set out in the contract.
  - b. Professional accountability. This involves the expectation that staff and administrators are answerable for their work as it affects others, for example in the use of resources, in their professional practice, or in programme outcomes.
  - c. External accountability. This is often motivated by the bureaucracy’s need to monitor the programmes which they are responsible for. The focus of evaluation is on whether the programmes are carried out on schedule, the use of resources and whether they produced the intended outcomes.
- ii. Evaluation for development. This type of evaluation, is carried out to improve the quality of an educational programme, usually an on-going one. It may be carried out co-operatively by an external evaluator and the insider staff. The guiding principles are the identification of strengths of a programme that can be built upon, and its weaknesses that should be eliminated.

- iii. Evaluation for curriculum development and betterment. Information from teachers and other relevant ELT professionals can be gathered over a period of time for this purpose of evaluation. Developments are monitored by the identification of strengths and weaknesses of the teaching and learning process.
- iv. Evaluation for teacher self-development. This includes raising the consciousness of teachers and other ELT professionals to what actually happens (as opposed to what is supposed to happen) in the language-teaching classroom. Evaluations for this purpose focus on the process rather than the end product (summative).
- v. Evaluation of teachers and teaching. This is to evaluate how well teachers conduct their teaching practice. Classroom observation provides insights on what teachers do rather than what they know. This purpose of evaluation is divided into:
  - a. Grading teachers. This includes reporting a grade, accountability and judgemental statements about teachers' classroom performance based on a checklist by an assessor. This type of evaluation is usually summative in nature.
  - b. Teacher development. Here is how the feedback from a formative classroom observation can be used to further develop or improve an aspect of classroom practice, or as part of curriculum betterment or teacher self-development.
  - c. Teacher self-development. This is a more participant-oriented evaluation through observation (e.g. microteaching) to raise teacher's awareness. The process can include self-evaluation, peer evaluation and collaborative group work.
  - d. Teacher self-evaluation. It is a process where teachers reflect upon what has taken place in the classroom with a view to improve their future performance.
- vi. Evaluation of learner outcomes. The implementation of appropriate tests to obtain feedback on the achievement and progress of learners.
- vii. Evaluation of teaching materials. This is done to see how well the teaching materials support the program. Materials evaluations are also done for making decisions to purchase them, or for textbook selection.

In contrast, assessment is a process centred on individual learner ability where it is carried out to gauge and enhance learner learning and to improve teachers' teaching as both learners and teachers are directly involved in it and mutually respond to it. Following are some of the purposes of assessments [1,3,8,17,18]:

- i. Assessment for pedagogic purposes (usually formative in nature):
  - a. To provide feedback for the benefit of the teaching and learning process.
  - b. To provide information on a learner's ability (strengths, weaknesses and needs).
  - c. To motivate and improve learners.
  - d. To reinforce teaching and learning
- ii. Assessment for classificatory purposes (usually summative in nature):

- a. For the selection or placement of a learner.
- b. To maintain standards
- c. To establish norms.
- d. To perform accountability functions.
- iii. Assessment for research purposes:
  - a. As part of an evaluation.
  - b. For experimental purposes.
  - c. To gauge knowledge or skills about language learning or language use.

## 6. THE MEANING OF 'EVALUATION' AND 'ASSESSMENT' IN ELT AND LANGUAGE PROGRAM EVALUATION

Based on the differences of the terms 'evaluation' and 'assessment' from the many perspectives that have been discussed, leading ELT experts have discerned the two terms primarily on the basis of the scope and purpose of each. Evaluation is accordingly acknowledged by leading ELT experts as the broader concept of the two in the sense that it is a systematic inquiry of the whole language program which is carried out for the purpose of making decisions or providing opportunity for reflection and action in which it can make use of assessment instruments as one of the forms of information gathering [8,14,16,20]. Brown, meanwhile looked at evaluation as "the systematic collection and analysis of all relevant information necessary to promote the improvement of the curriculum, and assess its effectiveness and efficiency, as well as the participants' attitudes within a context of particular institutions concerned". Norris, meanwhile mentioned the variety of program elements and purposes, and the improvement of program value by saying that evaluation is "the gathering of information about any of the variety of elements that constitute educational programs, for a variety of purposes that primarily include understanding, demonstrating, improving, and judging program value."

In other words, evaluation is not just answering how well students have done, but also addressing wider ranging questions such as how well the program has served the learners, educators and stakeholders, how much value for money the program has delivered, and how the program has fared in comparison to others or how effectively it has been executed in meeting its more eclectic objectives.

Assessment, on the other hand, is recognized as the narrower concept which is focused on the measuring of individual learner ability. It involves the implementation of tests and tasks given to learners to measure their individual performance, which also serves the purpose of the improvement and development of learner learning through a variety of ways. It is defined as "the range of procedures used to investigate aspects of individual language learning and ability, including the measurement of proficiency, diagnosis of needs, determination of achievement in relation to syllabus objectives and analysis of ability to perform specific tasks (p. 1)[8]." Norris looked at assessment as the systematic gathering of information about learner learning which is used in support of the teaching and learning process [11]. In congruence with this, Nunan said that assessment is a process of collecting information in making judgements about what individual learners can do in a target language [14]. Therefore, assessment is clearly distinct from evaluation,

where the latter involves valuing or judging the worth of a language program that is made up of a plethora of elements (of which assessment is one of them).

Having said that, it is also noted that assessment data can be part of the information required in an evaluation, or it can also be used for other purposes, such as the measurement of a learner's general ability, placement of the learner (in a particular area of work or a particular level of a course), to indicate a learner's progress (diagnostic purposes) and to provide feedback to learners for them to enhance their awareness and reflection [8].

## 7. CONCLUSION

In conclusion, to maintain a clear direction in the field of ELT in general, 'evaluation' is distinct from 'assessment' [1, 11]. It is the broader of the two concepts and applies to a wider range of issues involving the entirety of a language program. Assessment, conversely, focuses specifically on individual learner achievement. The assessment data can be part of the information required in a language programme evaluation or it can be used for other purposes. An accurate in-context understanding and usage of these two terms is important to reconcile ambiguity, and thereby facilitate a more efficient and effective sharing of knowledge in the collective quest for further progress in the field.

## REFERENCES

- [1] Brindley, G., D. 2003. Methodology. In Nunan, D. (ed.). Practical English language teaching, McGraw Hill, Boston, Chap. 15, pp. 309-328.
- [2] Brown, J.D. (1989). Language program evaluation: a synthesis of existing possibilities. In Johnson, R.K. (ed.). The second language curriculum, Cambridge University Press, Cambridge, Chap. 14, pp. 222-241.
- [3] Cohen, A.D. 2001. Second Language Assessment. In Celce-Murcia, M. (ed.). Teaching English as a second or foreign language, Heinle & Heinle, London, Unit V, Chap 35, pp. 515-534.
- [4] Garinger, D. (2002). Textbook selection for the ESL classroom. Center For Applied Linguistics Digest 02-10.
- [5] Harmer, J. 2007, 4th ed., The practice of English language teaching. Pearson, Essex.
- [6] Kiely, R. and Rea-Dickins, P. (2005). Program evaluation in language education. Palgrave Macmillan, Hampshire.
- [7] Lynch, B.K. (1996). Language Program Evaluation. Cambridge University Press, Cambridge.
- [8] Lynch, B.K. (2003). Language Assessment and Programme Evaluation. Edinburgh University Press, Edinburgh.
- [9] Merriam-Webster. (2017). Retrieved from <https://www.merriam-webster.com/>.
- [10] Nevo, D. 2009. Accountability and capacity building; can they live together? In Ryan, K.E. & Cousins, J.B. (eds.). The SAGE International Handbook of Educational Evaluation, Chap. 16, SAGE, Los Angeles, pp. 291-303.
- [11] Norris, J.M. (2016). Language Program Evaluation. The Modern Language Journal 100 (Supplement 2016): 168-189.
- [12] Norris, J.M. (2009). Understanding and Improving Language Education Through Program Evaluation: Introduction to the Special Issue. Language Teaching Research, 13(1): 7-13.
- [13] Norris, J.M. (2006). The why (and how) of assessing student outcomes in college foreign language programs. The Modern Language Journal 90(4): 576-583.
- [14] Nunan, D. (1991). Second language proficiency assessment and program evaluation. In Anivan, S. (ed.). Issues in Language Programme Evaluation in the 1990s. Singapore: SEAMEO Regional Language Centre, pp. 46-59.
- [15] Oxford Advanced Learner's Dictionary. (2017). Oxford University Press, Oxford. Retrieved from <http://www.oxfordlearnersdictionaries.com>.
- [16] Palmer, A. (1991). The role of language testing in language program evaluation. In Anivan, S. (ed.). Issues in Language Programme Evaluation in the 1990s. Singapore: SEAMEO Regional Language Centre, pp. 1-14.
- [17] Rea-Dickins, P. and Germaine, K. (1992). Evaluation. Oxford University Press, Oxford.
- [18] Rea Dickins, P. 2000. Classroom assessment. In Hedge, T. 2000. Teaching and learning in the language classroom. Oxford University Press, Oxford, Chap. 11, pp. 375-401.
- [19] Richards, J.C. (2001). Curriculum development in language teaching. Cambridge University Press, Cambridge.
- [20] Richards, J.C. and Rodgers, T.S. (2001). Approaches and Methods in Language Teaching. Cambridge University Press, Cambridge.
- [21] Stufflebeam, D.L. and Shinkfield, A.J. (1985). Systematic evaluation. Kluwer-Nijhoff Publishing, Boston.
- [22] Weir, C. and Roberts, J. (1994). Evaluation in ELT. Blackwell, Oxford.

---

\* For correspondence; Tel. + (60) 3-90513400, E-mail: [salim@upnm.edu.my](mailto:salim@upnm.edu.my)