Language Program Evaluation

Aurore Bargat, University of Illinois, United States

Abstract
An educational program is a formal learning experience in which students attend classes either in-person or virtually. Additionally, evaluation is the process of gathering and analyzing data to make effective decisions to improve a specific program (Ornstein & Hunkins, 2009). Putting both concepts together, we come to the idea of program evaluation, which refers to "the process of assessing the distinction or value of some aspect or the whole of a curriculum" (Sharma & Raval, 2019, p. 242). Overall, language instructors and language program directors need to regularly evaluate their language teaching methods, their materials, and the effectiveness of the language courses they are offering. Therefore, this presentation aims at defining program evaluation, reviewing the current literature (Tyler, Taba, Process, Stufflebeam, and Kirkpatrick models) and putting together a comprehensive tool to assess language programs effectively. Finally, the presenter will introduce an innovative evaluation tool that will provide key points to the audience to better analyze how well a specific language program achieves its educational objectives.

Keywords: Program Evaluation, Language Programs, Language Program Evaluation

iafor

The International Academic Forum
www.iafor.org

Introduction

This paper aims at defining program evaluation, reviewing the current literature, and putting together a comprehensive tool to assess language programs effectively. Such foreign language programs are currently being provided by public and private educational institutions where students are native English speakers. A relevant evaluation tool should discuss how well a specific language program achieves its educational objectives.

The term educational program can be defined as "the planned guided learning experience and intended learning outcomes formulated through a systematic reconstruction of knowledge and experiences under the auspices of the school for the learner's continuous and willful growth in academic, personal and social competence" (Tanner & Tanner as cited in Sharma & Raval, 2019, p. 240). An educational program is a formal learning experience in which students attend classes for several hours a day.

Additionally, according to Kiely and Rea-Dickins (2005), evaluation is a form of inquiry, which ranges from research to systematic approaches to decision-making. In short, it is the process of gathering and analyzing data to make effective decisions to improve a specific program (Ornstein & Hunkins, 2009).

Putting both concepts together reveals the idea of program evaluation, which refers to "the process of assessing the distinction or value of some aspect or the whole of a curriculum" (Sharma & Raval, 2019, p. 242). Therefore, as Rea-Dickins and Germaine (1992) mentioned, "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 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 program. And they must be evaluated" (p. 5).

Overall, "an assessment measure will be strongly related to educational experiences and unrelated to noneducational factors" (Pike, 2002, p. 140) such as gender, ethnicity, abilities, and how the assessment was administered.

For educational programs specifically, some of the most widely used evaluation tools are:

1. Tyler's Model

This four-step model was developed in the 1940s by Ralph Tyler. It has four steps:

a) Determining the program's objectives. According to Sharma and Raval (2019), this step is to "plan what do the students need to do to pass a level or to be successful in achieving a mastery" (p. 245).
b) Identifying educational experiences related to those objectives that a student has to go through.
c) Organizing those experiences "in a coherent way of a demonstration by teacher followed by an exercise by the students" (Sharma & Raval, 2019, p. 245).
d) Evaluating the objectives by having the teacher assess students and whether they have achieved the expected proficiency level or not (Sharma & Raval, 2019).

Some advantages of this model are its easy application and that it focuses on the program/course's strengths and weaknesses. On the other hand, several drawbacks are also present in this model, such as the lack of precise standards/rubrics, an increased focus on learning objectives, and not taking into consideration formative assessment (Guba & Lincoln, 1981).

2. Taba's Model

This model was created by Hilda Taba in the 1960s. She was an American curriculum theorist. Her model focuses on adequate development and implementation of a curriculum following an inductive approach. Taba emphasizes that the curriculum should be responsive to changes in the education sector and appropriate adjustments should be made accordingly. This model is very similar to the 1967 Wheeler's model, containing identical steps but in a slightly different order.

> a) Diagnosis of needs
> b) Formulating objectives
> c) Selecting content
> d) Organizing content
> e) Selecting Learning Experiences
> f) Organizing Learning Experiences
> g) Evaluation

This model focuses on the organization of learning materials and activities. It can be noticed that little to no consideration is given to the teacher's role.

3. The Process Model

This model was developed in the 1970s by Stenhouse Lawrence. He was an educational researcher focusing on curriculum development. He also founded the Centre for Applied Research in Education at the University of East Anglia (England). The Process Model focuses on teacher's activities and role as well as on student's activities. His ideas include a flexible, open, creative, and innovative curriculum as the foundation of an effective educational program. However, one of the disadvantages of this model is its lack of consideration for the learning environment and the context in which learning is occurring.

4. The CIPP Evaluation Model or Stufflebeam's Model

This model was created by Daniel Stufflebeam in the 1980s. According to Aziz, Mahmood, and Rehman (2018), this four-part model "can be effectively used for evaluating the quality of education" (p. 190).

a) Context Evaluation: assesses the needs and opportunities within a defined context or environment (Stufflebeam & Shinkfield, 2007). The objectives of context evaluation are to define, identify and address the needs of the target population, identify the problems, and assess if the goals are responsive to the desired needs or not (Stufflebeam, 2003). The different types of tools to use for context evaluation include surveys, document reviews, data analysis, and interviews (Stufflebeam, 2003).

b) Input Evaluation: determines the resources used to meet the goals of the program (Stufflebeam, 1983). Resources include budget, time, personnel, infrastructures, etc.

c) Process Evaluation: focuses on "the running of the program and teaching learning processes" (Aziz, Mahmood, & Rehman, 2018, p. 193). In this phase, implementation of decisions or corrective actions are taken (Patil & Kalekar, 2014).

d) Product Evaluation: does not focus on student's achievement but on "the skills, attitudes, knowledge, learning and abilities they attain which the student is going to use in life to benefit society" (Aziz, Mahmood, & Rehman, 2018, p. 194).

Overall, the CIPP model focuses on different aspects "during the beginning, implementation and designing of educational programs" (Aziz, Mahmood, & Rehman, 2018, p. 194). Then, outcomes are compared to objectives, weaknesses are reported, and changes are made to improve the quality of education (Sancer, Baturay, & Fadde, 2013).

The advantage of the CIPP Model is its emphasis on decision-making as an appropriate way for administrators to improve curricula. This model also has some drawbacks, such as the difficulties related to its implementation and its expensive maintenance (Guba & Lincoln, 1981).

5. Kirkpatrick's Four Levels Model

This model was developed by Donald Kirkpatrick in the 1990s with the purpose of assessing training effectiveness. It has four straightforward steps:

a) Reaction: Did the learners enjoy the course?
b) Learning: Did learning occur?
c) Behavior: Did the training change behavior?
d) Results: Did the training influence performance?

The answers to those close-ended questions would need detailed elaboration to ensure that adequate decisions are made to improve current language programs. Overall, this evaluation model might seem relatively minimal and potentially ineffective.

As demonstrated by the five models presented above, many challenges arise when it comes to evaluating a language program adequately and accurately. Pike (2002) notes that, when creating an evaluation tool, it is important to ask relevant questions about student learning and the processes used by teachers to facilitate this learning. For instance, for "how much" questions (such as level of satisfaction or amount of change), quantitative methods (such as surveys and exams) may be most appropriate (Pike, 2002). On the other hand, for "how" questions (such as how students' experiences affect their learning outcomes), qualitative methods (such as interviews and focus groups) may be most appropriate (Pike, 2002).

The 3x3 Evaluation Tool

It is the author's ambition to create a comprehensive and accurate evaluation tool for language programs. To start drafting such a tool, the previous literature review (as well as the work of Tufail & Embi in 2018) was extremely useful to help categorize the various aspects of a language program.

The author created the following evaluation tool (Figure 1), called 3x3, to best organize the nine categories considered key elements in the evaluation of a language program.
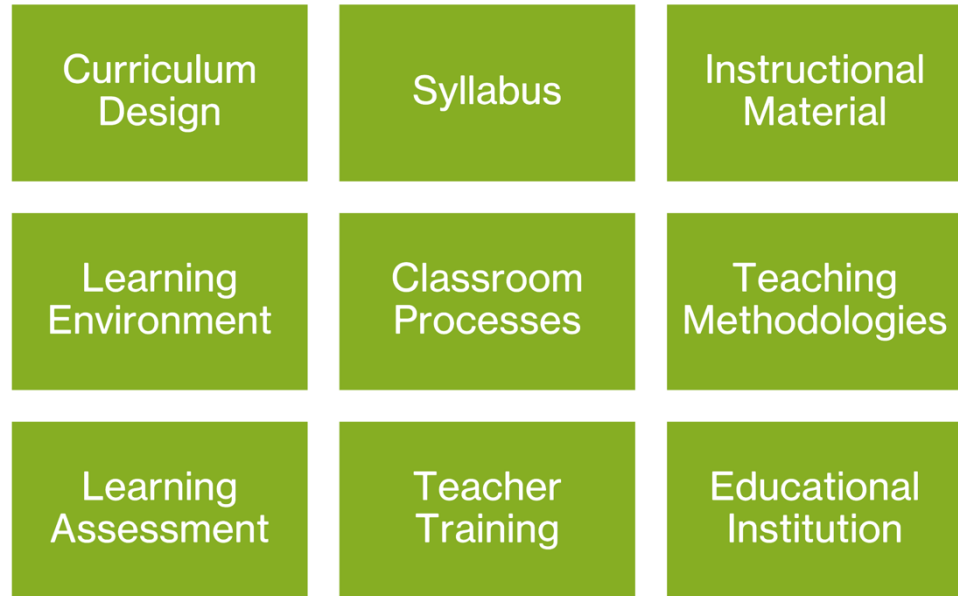
Figure 1: The 3x3 Evaluation Tool.

Consequently, the following explanatory tables (Table 1, Table 2, and Table 3) could be considered to organize the various aspects of a language program for a thorough and effective overall evaluation.

What is assessed	Why	Possible Reflections (non-exhaustive)
Curriculum Design	To provide insight into the program's organization	-Is the course/program F2F, hybrid, or fully virtual? -How are the units, chapters, and lessons organized? -What Learning Management System is used (if any)?
Syllabus	To assess how relevant the syllabus is	-What are the purpose and objectives of the course/program? -Is the syllabus bringing the students to the targeted level of language proficiency? -Is the syllabus fostering the students to practice all four modalities (speaking, reading, listening, writing)?
Instructional Material	To determine how relevant and effective the material is	-Are authentic reading and listening materials being used? -Is the target culture adequately represented and explained (going beyond stereotypes)? -Are clear objectives shared with the students? -Are all four modalities (speaking, reading, listening, writing) equally and regularly presented to the students? -Is a variety of students' assessments (formative, summative, etc.) often provided?

Table 1: Group 1 includes the areas of Curriculum Design, Syllabus, and nstructional Material.

What is assessed	Why	Possible Reflections (non-exhaustive)
Learning Environment	To evaluate the effectiveness of teachers aids and the relevant use of technology	-What tech tools are meaningfully used by the teacher? -Is the teacher using specific aids to support students learning? If so, what?
Classroom Processes	To provide insight in how the program is appropriately implemented	-What is the best environment for the course/program (F2F, hybrid, or fully virtual)? -How is the course/program made accessible for all types of learners (especially for students with reading or listening difficulties)?
Teaching Methodologies	To ensure the use of adequate teaching methodologies	-Does the teacher foster student-centeredness and student collaboration in the classroom? -Is the use of the target language fostered and used more than 90% of the class time by both teacher and students? -Does the teacher use various instructional strategies (PBL, IBL, TBL, etc.)? If so, which ones and how efficiently? -What assessment and feedback strategies are being used by the teacher?

Table 2: Group 2 includes the areas of Learning Environment, Classroom Processes, and Teaching Methodologies.

What is assessed	Why	Possible Reflections (non-exhaustive)
Learning Assessment	To ensure adequate assessment is used throughout the program	-What types of diagnostic assessments are available to the teacher(s) before the beginning of the language program/course? -What types of formative and summative assessments are used during the language program/course? -Are the students actual scores/results in line with the predictions?
Teacher Training	To assess if teachers are provided with sufficient professional development opportunities	-What professional development opportunities are provided within the institution (topics, lengths, etc.)? -What is the attendance rate? (If mandatory, is this new knowledge later implemented by the teachers?) -What professional development opportunities are recommended outside of the institution (and is time and/or financial support provided to attend)?

Educational Institution	To find out what administrative support and resources are available and if they are effective	-What resources are offered to the teachers by the institution? -What administrative support is offered to the teachers by the institution (easy report of teaching hours, printer access, etc.)? -Are the available resources used by the teacher effectively? If not, why?

Table 3: Group 3 includes the areas of Learning Assessment, Teacher Training, and Educational Institution.

For each category to be assessed, time should be set aside to reflect on the best way to set up data collection: Focus groups with teachers and administrators? Semi-structured interviews? Expert panels? Surveys?

For instance, to organize an effective focus group, a skilled moderator is needed to facilitate a 45 to 90-minute discussion between a maximum of six to ten people. Clear, pre-determined, and open-ended prompts are also needed (generally no more than 10). It is also best if someone takes notes on the participants' reflections to gain better insight into the topic at hand and for later reporting.

Similarly, when it comes to creating surveys for evaluation purposes, it is crucial to carefully select and create questions that would be useful in the evaluation process. Various options are possible: open-ended, multiple-choice, rank order, and rating scale. Each type of question has strengths and weaknesses that need to be analyzed before being used in the evaluation itself. This is to ensure that the evaluation is relevant and meaningful.

After collecting all necessary data, careful analysis and interpretation are also needed to provide concrete recommendations for implementation and potential corrective actions.

Conclusions

In this paper, a review of five different evaluation models for the classroom environment was performed and the draft of a comprehensive evaluation tool was created to adequately analyze every aspect of a language program.

As established by Tufail and Embi (2018), "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" (p. 174). This demonstrates that program evaluation is complex in nature and that the careful creation of a relevant and effective tool should be considered in order to adequately and accurately evaluate the effectiveness of a language program. For such programs, various aspects should be considered for evaluation, such as curriculum design, syllabus, classroom processes, instructional material, teacher training, the learning environment, and the institution itself.

References

Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP Model for Quality Evaluation at School Level: A Case Study. *Journal of Education and Educational Development*, *5*(1), 189-206.

Guba, E., & Lincoln, Y. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.

Kiely, R., & Rea-Dickins, P. (2005). *Program evaluation in language education*. Palgrave Macmillan, Hampshire.

Ornstein, A., & Hunkins, F. (2009). *Curriculum: Foundations, Principles, and Issues*. 5th Edition. Pearson.

Patil, Y., & Kalekar, S. (2014). CIPP Model for school evaluation. *Scholarly Research Journal for Humanity Science & English Language*, *2*(10).

Pike, G. R. (2002). Measurement issues in outcomes assessment. In Trudy Banta & Associates *Building a Scholarship of Assessment*. San Francisco: Jossey-Bass, pp. 131-147.

Rea-Dickins, P., & Germaine, K. (1992). *Evaluation*. Oxford University Press, Oxford.

Sancer, H., Baturay, H. M., & Fadde, P. (2013). Applying the context, input, process, product evaluation model for evaluation, research, and redesign of an online Master's Program. *The International Review of Research in Open and Distance Learning*, *28*(3), 274-290.

Sharma, H., & Raval, V. (2019). Curriculum Evaluation: Approaches and Models. *A Journal of Composition Theory*, *12*(10), 240-249.

Stufflebeam, D.L. (1983) The CIPP Model for Program Evaluation. In: Madaus, F.F., Scriven, M. and Stufflebeam D.L., Eds., *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Kluwer, Norwell, pp. 117-141.

Stufflebeam, D. L. (2003). The CIPP model for evaluation. In D. L. Stufflebeam & T. Kellaghan (Eds.), *The international handbook of educational evaluation*. Boston, MA: Kluwer Academic Publishers.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, & applications*. San Francisco, CA: Jossey-Bass.

Tufail, M. S., & Embi, M. A. (2018). Language Program Evaluation: Distinguishing 'Evaluation' from 'Assessment'. *Science International* (Lahore), *30*(2), 171-175.

Contact email: abargat2@illinois.edu