
Some Uses of Evaluation in Language Programs

Lyle F. Bachman

University of Illinois at Urbana-Champaign

While classroom teachers and program administrators often spend a great deal of time keeping records and gathering information for evaluation, it is not always clear that this information is used for evaluation, or what the use of the evaluation is. Information gathering is only one aspect of evaluation, and the other aspect -- decision-making, is often not adequately considered. The fundamental use of evaluation is to provide information for decision-making, and the uses of evaluation are thus related to the kinds of decisions to be made. The most lasting use of evaluation, however, is not the specific recommendations it yields, but rather the process itself, by which we can foster attitudes and processes that are at the heart of excellence in teaching.

Introduction

How many times have you said to yourself, "with all the records I have to keep and forms to fill out, I hardly have time to teach, let alone evaluate what and how I teach?" However, without a doubt, the demand to evaluate what we do plays a significant part in most language programs. Furthermore, this demand has become even greater in these days of expanding needs for language programs and static or dwindling resources, when we are being held increasingly accountable to local and state school boards, government funding agencies, boards of trustees and other clients, for what we are accomplishing.

Evaluation is seen by many as the key to such accountability. Over the years, however, both as a teacher and as a consultant, one of the most frequent complaints I have heard from classroom teachers and program administrators is that they have to spend too much of their time evaluating what they do. Certainly, what I often see in language programs is a great deal of information gathering (testing, interviewing, observing classes) and information storing, or record keeping. What is not so obvious, however, is whether this constitutes evaluation. The crucial questions to be asked here are, "What is done with all this information" and "What is the use of the evaluation?"

I'd like to share with you some thoughts about the nature of evaluation and some of the uses it can serve in language programs. I'll discuss two aspects of evaluation: information gathering and decision-making. Then I'll discuss two levels of decisions--about individuals and about the program itself--and the kinds of decisions that typically need to be made. Finally, I'll suggest that evaluation can be most useful if it is viewed as an on-going program itself, in which both the processes and the outcomes benefit the individuals involved.

But first, in order to illustrate some of the misuses evaluation can serve, let's eavesdrop on some "insider conversations."

Conversation 1 : "For the Record"

"I think we've got more data here than we can possibly handle, and it's still coming in."

"Never mind, it keeps people busy."

"But what'll we ever do with it all?"

"Who knows? We've gotta keep it all for our records, so enter it into the data base."

Conversation 2 : "Why Decide Today What We Can Decide Tomorrow?"

"Well, the committee on credentials has made its recommendations, but I'm afraid there'll be some objections if we make a decision now. Whadaya think we should do?"

"Well, we could always ask for an evaluation. That should take at least another year to complete."

Conversation 3 : "I Told You So."

"I Just know what he wants to do is all wrong. But nobody'll take our word for it. If only we could show the rest of the faculty how far off he is, nobody'd listen to him anymore, and we could get our proposal accepted. Any ideas?"

"Yeah. Why don't we ask for an evaluation? That'll prove that we were right all along."

Conversation 4 : "The Facts Leave Us No Alternative."

"You know our decision to cut the new program isn't gonna be very popular, especially since both the students and the teachers are pretty satisfied with it. But we all know it's too expensive to continue. How can we avoid getting blamed for dropping it?"

"Why don't we ask for an evaluation? That will prove that the program isn't cost effective."

Of course none of us has ever entertained thoughts such as these! Who would ever pile up information just to keep someone occupied? Or use an evaluation to delay making a decision, or to discredit someone, or to find a scapegoat? This is not what evaluation is for. Or is it? Such uses of evaluation are not uncommon. I have mentioned them not to brand myself as a cynic, but rather to provide some foils to underscore the productive uses of evaluation that I wish to discuss.

Evaluation

Let me begin with a definition :

Evaluation is the systematic gathering of information for the purpose of making decisions.

Included in this definition are two activities : information gathering and decision making. Let's talk about information gathering first.

Information Gathering

In any given situation, the chance of making a correct decision depends not only on the skill and experience of the decision-maker(s), but also on the quality of the information upon which the decision is based. Everything else being equal, the more reliable and relevant the information, the better the chance of making a correct decision. In order to assure that our information is both reliable and relevant, therefore, we must gather it systematically, according to a plan, rather than in a haphazard manner.

The information we collect typically falls into two broad categories : information related to *processes* in the program, and that which has to do with *outcomes* of the program. The outcomes that we are most frequently interested in examining are those stated as program objectives. In nearly every program, however, there are unanticipated outcomes -- both positive and negative.

My favorite example of an unexpected outcome is that of an individualized language arts program that I worked on in Thailand several years ago. One of the affective goals of this program was for students to learn self-direction in learning, and in terms of our evaluation, the program was very effective in achieving this objective. I'll never forget my chagrin, however, when the Deputy Minister of Education called me into his office to inform me that our "best" class of sixth graders had marched into their headmaster's office to protest the reassigning of their teacher to another class. We obviously had not designed our evaluation to collect information on the development of potentially antisocial behavior!

In evaluation we are interested not only in what the "products" of the program are, in terms of learning, satisfaction, and so forth, but also with how these outcomes come about. For this reason, we need to gather information on the processes that characterize the program. The problem here, of course, is that the very processes in which we are most interested-- learning and teaching-- cannot be directly observed. I would note that I consider teaching to consist of more than the actions we perform. Indeed, the most important aspect of teaching, I believe, takes place in the teacher's mind, in terms of both the conscious or unconscious choice of particular strategies and those spontaneous "bright ideas" that characterize creative teaching. The information we gather on these processes is indirect, and is typically based on techniques such as introspection and interviews or on the observation of activities that take place in the classroom. And in addition to the processes about which we do gather information, there are innumerable influences, both within the program and without, that we either do not or cannot observe, that determine the outcomes. Because of these limitations on the information we can gather, it is often difficult to attribute specific outcomes to specific processes, that is, to infer that a given process "caused" a given outcome. Thus, while we would like to determine how effective the processes of a program are in helping achieve its objectives, any statements we make in this regard are at best informed and reasoned judgements.

We typically gather both qualitative and quantitative information for evaluation. Verbal descriptions, ranging from performance profiles to letters of reference, as well as overall impressions, provide important information for evaluating individuals. So also may measures such as teacher ratings and test scores. Likewise, programs can be evaluated on the basis of qualitative information from narrative reports, anecdotal accounts or personal opinions, as well as quantitative data such as the number of hours in class, the number of graduates, or the percentage of students above a certain test score.

In order to avoid overlooking potentially useful information, our design for gathering information also needs to consider all the possible sources, such as students, teachers, administrators, parents, prospective employers, and members of the community. In other words, we need to be sure that we have obtained information from all relevant sources, which potentially include all persons affected by the program, either directly or indirectly.

Decision Making

A second activity in evaluation, as implied by the name, is that of assigning value, or of making value judgements. These value judgements, or decisions, are by necessity subjective. That is, although the information we gather may be objective, the process of sifting through that information, weighing the relative importance of the different pieces, considering how the various pieces are related, and so forth, involves applying a set of values and making subjective choices from among various alternatives on the basis of those values. And it involves not only the knowledge of what the various decision alternatives are, but also an understanding, or estimate, of what the consequences of these different choices may be.

Going back to our earlier discussion for a moment, one of the most frequent questions one hears in planning evaluations is "What kind of information do we need, and how much?" In order to answer these questions, we must consider the decisions that need to be made. We must keep in mind that the decisions we make are about *people*, and that these decisions will have some effect on their lives. It is therefore essential that the information we collect be both reliable and relevant, or valid. Very few of us, for example, would consider the opinions of friends and relatives reliable information for evaluating our students' language proficiency. Likewise, students' performance in other classes, such as mathematics or geography, would not be a valid basis for evaluating their performance in language class.

The *kind* of information we gather depends on the kinds of decisions that we need to make, while the *amount* of information we collect will depend upon the importance of the decision, or on the consequences of making a wrong decision. For example, if a classroom teacher must decide whether to proceed to the next lesson, he needs to know if the class has mastered the objectives of the current lesson. Depending on how well he has been monitoring the class's performance, it may not be necessary for him to gather additional information, such as giving a test, to make this decision. This is because the consequences of making a wrong decision are minimal. That is, if the teacher decides to proceed to the next lesson and finds that the class is not ready, he can fairly easily change course and return to the previous lesson for review. Consider, on the other hand, a situation in which test results are used to make nation-wide decisions about admissions to colleges and universities. In this case, the cost of admitting students who eventually fail and drop out may be enormous, both in terms of the educational resources expended and the misdirection of the students' time and efforts, not to mention the amount of personal loss involved. And the costs of denying admission to students who would succeed in an academic program are equally high, when we consider the potential loss to the educational institution and to society.

Uses of Evaluation

The fundamental use of evaluation in language programs is to provide information for making decisions that will improve the quality of the program. Conducting evaluation for any other purposes, as in the examples I gave earlier, constitutes a misuse of evaluation. In order to use evaluation productively we must subscribe to two basic assumptions:

- 1) Improvement in program effectiveness is both possible and desirable.
- 2) Information on processes and outcomes is essential to improving the effectiveness of the program.

Only those who feel that language instruction is an entirely futile activity would, I believe, question the first assumption. And even in situations where the evaluation may suggest that a program should be terminated, this might be viewed as an improvement in the educational system as a whole if that program is failing to meet its objectives, or if the costs of the program outweigh its benefits. If we accept the second assumption, it means that we consider feedback an essential mechanism for the continued effectiveness of the program. Without this assumption there is no reason to evaluate, since there are no decisions to be made. Most of us do accept this assumption, since we constantly conduct evaluation, whether we call it that or not. Any time we revise a lesson, change an approach, or alter a procedure on the basis of our experience with a group of students, we have used evaluation.

Levels of Decisions

The decisions that we typically make in language programs operate at two levels: the level of the individual, which we might call micro-evaluation, and that of the program itself,

which we might call macro-evaluation. The most frequently evaluated individuals in any language program are the students. The first decision that may be made about students is whether or not they should enter the program. In general educational programs, such as primary schools, entrance is nearly automatic with age, while in language programs, some evaluation of readiness, language dominance, or language proficiency may be required.

Many language programs include classes for different levels of ability, and for different needs and interests. In such programs, therefore, decisions regarding the placement of students into appropriate classes must be made. These decisions may be based on a variety of information, including native language, area of interest, amount of prior instruction, and scores on language aptitude and language proficiency tests.

Feedback on the effectiveness of student learning is generally of interest to both teachers and students. This information is useful to the teacher for making decisions regarding the speed and sequence of instructional activities. For example, if new material is being introduced at an inappropriate rate, either so slowly that the students lose interest, or so quickly that students cannot comprehend it, the effectiveness of learning is likely to decrease. Teachers also need to make decisions about when to move on to another unit of instruction. Although such decisions are typically made on the basis of qualitative feedback, some programs include short tests at the end of each instructional unit for this purpose. One decision that is seldom based solely on qualitative information is the evaluation of student achievement for assigning grades. In most cases grades are assigned on the basis of performance on tests in addition to classroom performance. In a typical program, for example, the final course grade might be determined by a weighted composite of mid-term and final examinations, class quizzes and teacher rating of classroom performance.

Decisions are also made about other individuals in the program. Teachers are hired, assigned to classes, promoted and so forth. Somewhat less frequent, but equally important, are decisions about administrators' effectiveness. Finally, policy making individuals, such as school boards and boards of trustees are themselves evaluated periodically by the electorate.

We also make decisions about programs. In revising an existing program, for example, we will want to evaluate specific components in terms of their appropriateness to the needs and interests of students and teachers, their effectiveness in achieving the objectives of the program, and their efficiency in utilizing program resources such as personnel, time, space and equipment.

Types of Decisions

The decisions we make can be classified according to the time at which they are made and their impact. *Formative* decisions are made during an individual's participation in the program or, at the program level, during the development or revision of a program. Their impact is to diagnose weaknesses and improve performance, of either the individual or of the program itself. *Summative* decisions, on the other hand, are made at the end of the individual's participation in the program, or at the completion of program development. They involve the acceptance or nonacceptance of performance, of either the individual or of the program. At the individual level summative decisions generally take the form of a grade or a pass/fail decision, while at the program level, they typically involve the adoption or non adoption, continuation or discontinuation of the program.

Evaluation as a Program

Now I would like to discuss evaluation as a program in itself, a program that can be examined in terms of processes and outcomes, and which is itself subject to evaluation. As with any educational program, evaluation has a stated objective: to provide information for

making decisions that will improve the effectiveness of the program. In evaluating the effectiveness of program of evaluation, we would need to gather information on both processes and outcomes. Let's begin with outcomes, since they are the most salient.

Outcomes

The outcomes of an evaluation will generally include a body of information and a set of recommendations for action. The question, in evaluating the effectiveness of the evaluation, is how useful the information is and whether implementing the recommendations does, in fact, improve the effectiveness of the program. If the evaluation has been carefully conducted, the information gathered can identify specific areas of strength and weakness in teaching materials, providing guidance for writers in revising them. It can also identify innovations in teaching that may have arisen spontaneously in the classroom and that might have gone unnoticed without a program of evaluation. Information obtained about students can be useful for modifying or expanding the program to better meet changing needs and the needs of other client groups. Another area in which evaluation information can be useful is in the efficient allocation of resources. It is not uncommon to find programs that overallocate resources in some areas and underallocate in others, and that could improve their overall efficiency by shifting resources around.

In evaluating the recommendations, one criterion might be whether or not they are adopted. This, of course depends not only upon the quality and feasibility of the recommendations themselves, but also upon other factors that may influence decisions, such as social and political pressures, the personalities of the decision makers, and the educational history of the system. A better criterion, assuming that the recommendations are implemented, is the extent to which they improve the program. And this, of course, leads into the next cycle in the evaluation.

Processes

The processes that are part of an evaluation are less obvious, perhaps, than the outcomes, but I believe they can be infinitely more useful in the long term. The process of identifying decision alternatives, developing a plan for gathering reliable information that is relevant to these decision alternatives, gathering that information from a wide range of sources, and then utilizing that information to reach decisions can have several positive effects. One result can be a group of individuals who are better able to participate in and make use of evaluation information. This is extremely important, since it strips evaluation of its mystique and reveals it as a process that is not only worth doing but doable. Gathering information from sources outside the program itself can also serve to inform the public at large about the values, objectives and activities of the program and to foster closer relations among program staff, parents and community leaders. Finally, the assumptions, or beliefs, if you will, upon which a program of evaluation are based -- that program effectiveness is possible and desirable and that it can be best achieved through feedback about processes and outcomes -- can foster an attitude of openness to innovation.

Let me try to clarify this view of evaluation with an example of two different approaches to conducting evaluation. In one approach, the individuals responsible for the program contact an outside evaluation "expert" and ask him or her to evaluate the program. This evaluator, after an initial consultation with the program staff, typically gathers information, analyzes it, and presents those responsible for the program with a set of recommendations, which may consist of a number of decision alternatives. The "outcomes" of this process thus consist of information and a set of recommendations. The usefulness of these outcomes will depend on how sensitive the evaluator has been to the needs of the program staff and upon the relationship between these two parties. What is not likely to result from this approach is an increased capability on the part of the program

staff to continue evaluation and to use its findings appropriately. Furthermore, individuals in the program may develop a negative perspective on evaluation and, perhaps, to the idea of improvement through feedback and innovation, especially if the evaluation is unfavorable, or if they disagree with the recommendations.

A quite different approach is for the evaluation to be conducted inhouse by the program staff, either alone or in cooperation with an outside evaluator who assumes the role of consultant. In this approach, the program staff is directly involved in every phase of the evaluation: determining the questions to be addressed, designing the gathering of information, analyzing the information, and interpreting the results in order to reach conclusions about the most beneficial course of action. This approach can yield the same outcomes as the first. In addition, because the individuals in the program have been part of the process of the evaluation, they will emerge from the experience with a better understanding of the considerations, problems and uses of evaluation, and will be better able to utilize the outcomes of this particular evaluation for their intended purpose, the improvement of teaching and learning.

Conclusion

In this paper I have discussed some considerations to be made in conducting evaluation and some of the uses of evaluation in language programs. I have described evaluation as comprising two activities: gathering information and making decisions. The types and amount of information we need to gather depends upon the kinds of decisions to be made and their importance. Decisions operate at two levels: that of the individual and that of the program itself. Decisions are also of two types: those aimed at improving performance (formative) and those that constitute the acceptance or nonacceptance of performance (summative).

We often view evaluation in terms of its outcomes--the information provided and recommendations made. It is my belief, however, that the *process* of evaluation is of more lasting usefulness and benefit to a program and the individuals in it than the specific information or recommendations any given evaluation may yield. That is, by viewing evaluation itself as an on-going program, we can foster the attitudes and processes that lie at the heart of excellence in teaching.

The Author

Lyle F. Bachman is Associate Professor of English as an International Language and Director of Overseas Projects at the University of Illinois at Urbana-Champaign. He teaches graduate courses and has published widely in the areas of language testing, English for specific purposes, curriculum development and evaluation. His publications include *Reading English Discourse* (Prentice-Hall), and *Fundamental Considerations in Language Testing* (forthcoming, Addison-Wesley).