

A PHILOSOPHY OF EVALUATION

Definitions

Evaluation can be divided into three categories. These might be termed **progress evaluation**, **diagnostic evaluation** and **proficiency evaluation**. I think of a progress evaluation as part of a basic program of language learning. An example would be periodic exams over material covered in a language course. Progress evaluations might also be functional, testing whether the learner can handle a certain situation or fulfill a certain assignment in a real-life situation.

Progress evaluation in the Standard Curriculum involves the completion of the basic **components**. Using the checklists provided here, or similar ones based on this design, the learner and the agency have a clear and objective measure of the learner's progress. Completion of the course entails fulfillment of each activity and completion of each component on the checklists. The Progress Report form provides a thorough self-evaluation, which is objectified when the form is sent in to an agency official as a report.

A diagnostic evaluation is skill oriented with a goal of discovering deficiencies or weaknesses which would then be dealt with. A Refresher Course might begin with a diagnostic evaluation, then a learning plan for the course could be developed emphasizing the needs of the learner.

A proficiency evaluation is not concerned so much with certain material or specific grammatical features or vocabulary, as with overall combined skills and general facility in the language. Progress evaluations test mastery of specific material studied. A proficiency evaluation, on the other hand, is not limited to any specific material, but covers the whole range of the language as spoken or written by native speakers of that language.

Purpose

Evaluation at any level can reward effort and accelerate progress. Evaluations provide an opportunity for review and enable the learner to consolidate what has been learned up to that point. Evaluations enable the learner and tutors or language helpers to discover or clarify areas of weakness which can then be strengthened. Evaluations can also provide the basis for guidelines for continuing progress.

PROFICIENCY EVALUATION

Concepts

A proficiency evaluation is different from a test over specific material studied in a course. The instruments should be designed to probe a person's full range of skills in reference to the total range possible. In contrasting a "test" and a proficiency evaluation, an analogy may be drawn from the American academic system.

In a particular university course, a student is tested over the specific material taught in the course. A basic language course is similar. There are other special examinations which examine overall knowledge in a certain discipline or combination of disciplines. The Undergraduate Record Examinations and the Graduate Record Examinations are examples of the latter.

The URE in philosophy, for instance, which is taken by university seniors in philosophy, does not examine on the basis of material studied in a particular university course, but includes questions from the full range of Western philosophy. **It is not possible for a candidate to know everything that might be asked** on a URE. But the marking of the exam shows where the candidate falls in the range of **total possible marks** in an overall performance. Such an exam necessarily shows not only what the candidate knows, but also what the candidate does not know.

The GRE, likewise, is designed to show how much of the possible material has been learned, or how much skill has been attained, in reference to how much **might have been attained**. A candidate may have learned more or gained less, but a particular school or department sets a minimum qualification for its own purposes.

Professional qualifying exams fall into this category also, such as examinations for the bar or government radio engineering licenses. The proficiency evaluation in language considers the total accumulated skills compared to the total range possible in the language.

A test evaluates the candidate **as a student** of the language. A proficiency evaluation evaluates the candidate **as a speaker** of the language. (That is, how does this particular candidate compare to a native speaker of the target language.) Thus it is important to make a technical distinction between tests and evaluations.

In implementing evaluation, the overall value to be gained for the agency as a whole, including the individual's productivity and success, must be considered paramount. The mild intimidation some people feel in face of **any kind** of evaluation is minimal compared to the increased sense of worth and self-esteem most gain from the success of fulfilling the evaluation or getting the help and improvement in communication stemming from it.

The Instruments

Technical and practical factors involved in evaluating language proficiency and cultural awareness have been taken into account in the development of the formats and instruments. The evaluation instruments have been prepared in such a way that it should be possible to photocopy them for use as they appear. The original A-4 size master copies have been reduced. They can therefore be expanded to regular size for photocopying or modification.

Factors for Evaluation

The definition of proficiency levels used here was originally developed by the Foreign Service Institute of the United States Department of State. The FSI is the general standard for language evaluation. The evaluation format based on the FSI definitions has been used by the Peace Corps in its training programs around the world.

A major deficiency I find with the FSI/Peace Corps approach to evaluation is that only technical language features were considered in evaluation of proficiency levels. The only factors considered were grammar, vocabulary, fluency, pronunciation and comprehension. This overlooks the cultural and social aspects of communication.

The instruments presented here evaluate the combined skills of communication, including cultural awareness. In the oral evaluation, only 50 points out of 100 are given to these technical linguistic features. Factors to be considered in the cultural and social skills in an African setting, for instance, would be:

Ability to relate to the Africans, sensitivity to African patterns of organization, decision-making, ways of learning, etc.)

Proficiency Level Checklist

The checklist is based on the FSI proficiency definitions. The Brewsters improved on the FSI checklist in their book *LAMP*. The one presented here is more detailed and specific. Some might feel that the fact that it is therefore longer might be a disadvantage, but this should actually make it a more effective and easier-to-use tool for making a **more precise evaluation** of the actual proficiency attained.

In addition, more attention has been given to cultural factors, specific linguistic features and ministry activities. I have also incorporated Dr. Donald Larson's definitions for independent learning and cultural involvement.

In the first four levels the sequence of areas evaluated are:

- Social situations and topics
- Church and ministry skills
- Language learning skills
- Language features and usage

The definition of each proficiency level has been worded so as to be applicable to every society and most languages. The language-feature descriptions show some bias towards the Bantu language family, the largest in Africa, where these instruments were first used. The amount of time and training required to reach a certain level will vary from language to language, as will the specific linguistic features. Nevertheless, a person with Level 4 in both Zulu and Tswana, or Gujarati and Punjabi, for example, should have approximately equal linguistic competence in the two languages.

Note that for languages like Amharic or Thai, literacy skills may require more time, in comparison to oral skills, than that indicated in these levels. Following principles presented in Section I of the book, **priority should be given to oral skills**. Literacy can easily be gained after the language is minimally learned.

All the ratings except level 5 may be modified by a plus [+] indicating that proficiency substantially exceeds the minimum requirements for the level involved, but falls short of those for the next higher level. Level 5 is native/bilingual proficiency. Though no agency assignment would normally carry the requirement of Level 5, some missionaries would attain this level if working in a stimulating atmosphere and with strong motivation.

The checklist should be used in the following ways:

1. By the **learner** to evaluate personal progress in language skills and cultural awareness at any and every stage of development.
2. By language **teachers or coordinators** to evaluate progress **during** an initial language course, or at stages in self-directed learning by the learner.
3. By teachers or coordinators at the **completion** of an initial course or other specific periods of residential study.

4. As a **diagnostic tool** at the beginning of a Review or **Refresher Course**.
5. As part of any **Proficiency Evaluation**.

The checklist can be used as is by instructors, language helpers and language coordinators, or it can be modified by deleting or changing the first-person references. The "plus" levels are unnecessary for the evaluations, as specifics are in focus. A total score for skills on this checklist, along with general cultural awareness shown, should be entered on the score sheet for "Evaluation of Communication Skills, Oral Skills," in the sections labeled **Social Skills** and **Cultural Awareness**.

Format for Oral Proficiency Evaluations

The traditional format for a proficiency evaluation has been an interview-conversation with the evaluators. This normally needs to be about one hour long. Sometimes a shorter period is sufficient for the learner to demonstrate his/her highest level of proficiency adequately. Every description I have seen for language proficiency evaluation proposes this format.

We have also used **two other formats**, which are actually **variations** of one alternative approach for evaluating proficiency. The first involves a one hour period of time also, but is in a **free-form conversation** format. For instance, we have had evaluators walk around the work site, such as a school campus, with an individual, or a couple. The evaluators make notes, as necessary, and everyone knows this is an evaluation. But the atmosphere is usually more relaxed, the learner is in a familiar setting, and more "in command" of the situation. This diminishes or eliminates the element of intimidation, or anxiety.

Such a discussion may involve a tour and explanation of the work situation, introduction to coworkers, explanation of procedures, etc. Often areas on the checklist, or similar topics, will come out in the course of such a conversation. Evaluators are free to introduce any topics they wish to, and may also take the lead at times, in order to "check out" certain areas.

A **second alternative** is to evaluate learners over a period of time while teaching or assisting them in various learning activities or simply observing them in various situations. The evaluator would spend some time in the individual's home, accompanying the person on trips to the market, business offices, churches, schools, etc. The time for this might vary. At the end of the period of time, the evaluator would fill out the evaluation instruments. Such an evaluation should be conducted by two or more evaluators. The evaluation might be given in conjunction with some formal learning sessions or drills.

Another separate period of time might be spent similarly with the learner by another evaluator, who would perform a similar summary evaluation. A supplementary conversation might supplement such long-term evaluation procedure.

While many learners prefer the informal oral evaluation, some choose the interview format when we offer them a choice. They may feel it is more trouble to arrange field observations in the area of work than to come to the language center for a conversation with evaluators.

Panel. The panel should consist of no fewer than two, preferably three, national speakers of the language, preferably native speakers. (In the multilingual setting of East Africa, for instance, where native speakers of Swahili are in a decided minority, though Swahili is the primary language for many educated nationals, who would serve adequately as evaluators of a foreigner's Swahili.) There is some difference of opinion among "experts" on whether foreigners should be involved. Some candidates have felt it adds an element of support to have an experienced colleague participate. Some feel intimidated. I have found that most express no opinion.

Interview. The conversation should cover as wide a range of topics as possible. Discussion should be conducted in such a way that the panel has a chance to observe the speaker's flexibility in structure and vocabulary within critical topics, as well as background awareness of cultural factors and experience in society. The first portion of the

interview/conversation might be done with only one person, while evaluators observe. Each panel member, however, should have a chance to talk with the learner, in order to evaluate flexibility of comprehension.

In addition, a story segment can be included to help demonstrate oral comprehension and breadth of vocabulary. Such stories might be chosen variously to fit the various levels of proficiency for the learner's stated level for job assignment. At the basic level, the story should be read to the learner and oral comprehension questions asked by one member of the panel. Discussion of the topic or events in the story could be included. Higher-level stories could be read in advance by the interviewee, with questions and discussion done orally as part of the interview.

If no field observation is involved, an optional section of an interview type evaluation would be for the interviewee to teach a lesson sample related to his/her area of work or expertise, such as a lesson on music, a Bible lesson, TEE lesson, a preventive health lesson such as might be taught in a public health clinic, etc.

In summary the oral evaluation should include the following components, whether in an interview or in a field work situation:

1. Some basic conversation with an individual or group on general topics, while panel observes.
2. Conversation with panel members on topics from the proficiency level checklist.
3. Optional presentation of a prepared sample lesson, or something work-related.
4. An oral comprehension exercise: a story read by a national, with some basic comprehension questions and discussion of the story.
5. An oral production exercise might be included: a portion of scripture or a newspaper article or story to be read by the learner.

Instruments. The two instruments used are "Evaluation of Communication Skills, Oral Skills" and the "Proficiency Level Checklist." Points are given only on the items on the "Evaluation" sheet. Topics covered in the conversation should be checked "yes" or "no" on the checklist, with appropriate comments. Points for Topical Skills and cultural knowledge are given on the "Evaluation" scoring sheet, under **Social Skills** and **Cultural Awareness**. (See below.)

The evaluators should use the oral evaluation instrument to give a score for the performance of the candidate. The technical points are well-known: Oral **Comprehension** (hearing and understanding), **Grammar** Correctness, Adequacy of **Vocabulary**, Correctness of **Pronunciation**, and **Fluency**. Top scores represent skills of a native or bilingual speaker of the target language.

The basic goal is **communication**. Therefore other values in addition to technical skills should be considered. Some people can get their idea across and effectively communicate, even with obvious deficiencies in technical points, such as grammar or pronunciation. Evaluate this ability in the category of **Understandability**. This will be weighted to 20%.

Under **Creativity**, consider how **flexible** the person is with the language, how great a variety of forms and structures he/she can use, how many ways he/she knows to say the same thing in subtly different ways. If the person can make only one or two set statements about a certain topic, low marks should be given for this skill. It is important for the conversation to cover many angles of each important question, in order to evaluate creativity.

Under **Social Skills**, give marks for the topics in the Proficiency Level Checklist. Observations should be made in field situations, in role-play or in discussion of certain situations, about the person's approaches, interaction, and awareness of social protocol. Watch also for body language, use of idioms, intonations, etc.

Under **Cultural Awareness**, watch for insights the person has on the culture in the way he/she speaks about experiences, the people, history, wedding or burial customs, proverbs, etc., as well as attitudes towards traditional concepts or national attitudes. This does not necessarily imply agreement or acceptance, but sympathetic awareness, and even areas of conflict between the local culture and the candidate's home culture and heritage.

Problems. As we have used these materials for several years in our Swahili program, I have found that the African evaluators tend to gradually mark higher and higher. I am almost certain from observation that this relates to the distinction made above in the philosophy of evaluation.

Our evaluators find it difficult to really judge foreigners against native speakers. They continue to evaluate candidates as students. Thus they evaluate on the basis of what could be expected of most learners, not completely on the way a Swahili actually speaks his own language. It is easy to think, "She is doing very well **for a foreigner.**" While this is good to know, the ultimate goal is to do very well in comparison to the Swahilis. This is also affected by cultural factors.

Evaluating Literate Skills

In some societies, literate skills may not be of great importance. In others, there is a literate tradition, but the writing system makes it difficult to attain high literate skills early. In others, literacy may be simply expected, and must be attained by a foreigner who wishes to be accepted and respected. Priority should be given to oral skills from the first, due to the personal relationship and social factors. The agency will then need to determine the importance of literate skills in reference to specific job assignments.

Levels 1, 2 and 3. General literate skills should be evaluated through use of a selection of written exercises, as outlined below. Sections 1 through 5 of the exercises should be used for evaluation of levels 1, 2, and 3. To qualify for level 4 or 5, sections 6 and 7 must also be attempted. The actual exercises used can vary, and the actual points in each section, but the **proportion** of skill areas should be the same.

The exercises should be marked (corrected) on the paper and the score transferred to the score sheet for easy interpretation into proficiency levels. These questions and comprehension selections should be prepared by nationals (or skilled expatriates) familiar with testing and marking where possible.

Note that it is possible to make a written evaluation too easy. Be sure the exercises and questions are truly representative. See the chart below the score sheet. For instance, note that the levels indicate harder marking than a final exam at the end of a six-months course.

I have observed the same problem with written evaluations as mentioned above for oral evaluations. It is too easy to correct the proficiency evaluation exercises as a student's exam in a course. The written exercises should be judged by expectations for a native speaker of the target language.

1. Grammar, vocabulary, sentence structure questions. (Negative-positive; passive-causative, conditional-future-past tenses, person subjects-objects, etc.): 40 points
2. Essay (marking includes grammar and vocabulary, usage): 20 points
3. Religious reading comprehension: questions and discussion: 20 points
4. Secular, traditional reading comprehension: 20 points
5. Idioms, usages and proverbs: 20 points

(Total Points: 120)

Level 4. To qualify for level 4, a candidate should also sit for these two additional sections, reflecting the high-level skills required for this level. The table following the score sheet for the level of proficiency indicates that those wishing to qualify for levels 1 through 3 may simply leave off these additional exercises.

These exercises will need to be marked by a university professor, national editor-publisher, writer or similarly qualified person. Your government examination commission or university may offer such services to agencies such as missions. (These two exercises would not have to be marked by the same person marking the exercises in the level 1, 2, and 3 qualification section.)

1. Editing skills: A book, article or magazine selection with mistakes to locate and edit to proper or better form or style. (Corrupt a lesson from a Sunday School quarterly, an article from a newspaper or a page from a book. Or likely you can find one ready-made!) 25 points
2. Advanced comprehension: A selection from a journal, secondary exam materials, or similar, with questions. This should be on a topic that will evaluate high-level vocabulary and usages. 25 points

Total for all written exercises: 170

Combined Scoring

With the instruments and formats provided, you will have a score giving the proficiency level for two different categories of communication skills. Oral Skills and Literate Skills will be evaluated separately. This would mean that it is possible for a person to have different proficiency levels for the two categories of skills.

In such a case it would be up to the agency to determine the appropriate action if the person demonstrates a lower level than required in one category but an acceptable level in the other category. The separate categories allow some flexibility in the emphasis given to oral vs. literate skills.

It may be desirable to combine the scores for oral and written skills for a single score for overall proficiency. This also provides some flexibility in determining proficiency and follow-up action, by allowing the stronger category of skills to balance the weaker category through the combined weighting. (Note that literate skills become more important the higher the level of proficiency.)

Use the following scale to determine proficiency based on the total of both categories of language skills (a total of 270 points):

Score	Level	Score	Level	Score	Level
16-60	0+	61-110	1	111-170	2
171-214	3	215-255	4	256-270	5

(Approximate percentage values of the levels, based on 270 points, are as follows: 0+=6-22% 1=23-41% 2=42-63% 3=64-79% 4=80-94% 5=95-100%)

REPORTING

At the end of the initial intensive period of study in the language (whether in residential study or self-directed), a review exam and evaluation should be given. The **Checklist** can be used at this time, or even the other instruments. The Oral and Literate Skill Evaluation will be of more value at the end of a term, evaluating cumulative skills.

A Proficiency Evaluation should be planned during the first term of service to determine the level attained. Some have qualified at Level 3 before two years. At least one person I know attained Level 4 in an African language in less than two years. But the required level of proficiency, orally, should be minimally attained by no later than 42 months on the field. High literate skills may take longer than the same level in oral skills.

The original copies of the evaluation instruments, including the checklist marked by each evaluator, should be given to the learner evaluated. Photocopies may be kept in the individual's permanent file.

Now -- what do you do with the results of the Proficiency Evaluation? The level of proficiency would be one factor in considering the assignment of the individual, along with relational skills, and professional skills. There are basically three possibilities in interpreting the results.

(1) The evaluation may affirm that this individual has attained or exceeded the level of proficiency required for the job assignment. The appropriate recommendation in this case would be to affirm the person in the same assignment or any assignment up to the demonstrated level of proficiency.

The evaluation might show up specific deficiencies in the individual's communication skill. This might involve a second and third alternative.

- (2) Suggestions for overcoming the deficiency, but with no "official" requirements.
- (3) Specified program for removing deficiencies.

In the latter case, there might be two alternatives again for implementing this.

- (a) Design a study plan to be carried out personally before furlough or upon return from furlough.
- (b) Specify a period of study at a language center.

A standard procedure should be established for making regular reports on the progress of missionaries in language proficiency and cultural awareness. Records should be kept in permanent files for future reference. In this way a diagnostic history will be on file to assist in ongoing support for each individual. The sponsoring or supervising agency should establish procedures detailing when reports are to be given and what reports are to be kept in the permanent record for each individual.

This is an updated excerpt from the book *Planning and Evaluating Missionary Language Learning*, by Dr. Orville Boyd Jenkins (Limuru, Kenya: Communication Press, 1989), **SECTION IV, EVALUATION**

Copyright © Orville Boyd Jenkins 1989, 2007

Permission granted for free download and transmission for personal or educational use. Other rights reserved.