Program Evaluation in College Foreign Language Education: Need, Capacity, Development

I. Introduction and Project Overview

College FL educators are faced with the sometimes daunting task of evaluating their programs, not only in response to accreditation and accountability pressures, but also in order to understand, improve, and ensure the quality and value of their educational efforts. The basic mission of the Foreign Language Program Evaluation Project is to help FL educators make the most of evaluation processes and products as they engage with a variety of program evaluation demands. In order to do so, we seek to understand the ‘why’ of evaluation—and the settings and constraints of FL programs—first and foremost. Based on a sound understanding of the actual purposes and contexts for evaluation, we then develop and disseminate a range of strategies, models, templates, and tools that are tailored to priority needs of real language programs. Finally, we investigate, revise, and improve these capacity-building strategies and resources through case studies of their meaningfulness and usefulness in representative college FL programs.

This report: A foundational purpose of this project, then, is to investigate and articulate a comprehensive understanding of program evaluation needs in the actual contexts of college FL programs across the United States. The current report presents initial findings from the needs analysis (Phase I), which was based on interviews and nationwide surveys of FL program administrators.
II. Interview Study and Survey Development

1. Overview
In order to generate initial understandings about current evaluation issues in college FL education, interviews were conducted with informants at several U.S. institutions. Interview questions were based on a review of the language program evaluation literature and the outcomes of focus group sessions at the 2005 ACTFL convention. Interviews were conducted in person or on the telephone and lasted between 30 and 45 minutes each. Data were transcribed and analyzed by three coders to extract (a) the ways in which FL educators talk about evaluation, and (b) major themes associated with FL program evaluation. These findings provided the basis for development of a survey that would address the program evaluation themes of interest to FL educators in the language that they use to discuss evaluation.

A. Purpose
(a) Gain a general idea of the range of practices, needs, and concerns of FL educators in evaluating their programs
(b) Provide an empirical basis for developing the large-scale survey of FL educators across U.S. colleges.

B. Participants (N = 21)
- WHO: Chairs, directors, section/language coordinators, assessment specialists
- FROM: Hawaii and mainland U.S. colleges, public and private, small and large, 2-year/4-year

C. Topics
(a) Participants’ understandings of the meaning of program evaluation
(b) Characteristics of participants’ language programs
(c) Current and desired evaluation practices in the program (impetus/purpose/methods/use)
(d) Current status and needs for capacity development to conduct program evaluation
(e) Issues and concerns regarding program evaluation

D. Findings
- Considerable demand → wide range of practices & capacities → substantial setting constraints → desire to resolve and improve evaluation practices → key suggestions for resource types that would help in doing so
- Findings indicated major thematic areas for further investigation, as well as range of possible items in each theme

2. Evaluation themes

III. Survey on college foreign language program evaluation

Participants: The current membership data base of the Association of Departments of Foreign Languages was utilized as the population frame for this survey. Members of the ADFL consist of chairs, directors, supervisors, coordinators, and others involved in the administration of college FL programs across the U.S. Only current members were contacted in order to maintain a focus on evaluation issues of immediate priority.

Survey instrument: Based on themes that emerged from interviews of FL program informants, a web-based survey instrument was developed. Following questions about FL program contexts, respondents were presented with the following definition of program evaluation, in order to ensure a common understanding of this key construct.

"Evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming." (p. 23)


Seven survey sections then covered evaluation impetuses, uses, focus, methods, capacity, needed resources, and concerns. Respondents indicated their answers on Likert-style rating scales, and each section also sought open-ended comments. The online survey was pilot-tested by 12 foreign language educators in diverse program settings, and survey format and items were substantially revised to increase comprehensibility and ease of completion.

Survey procedures: An initial email invitation was sent to participants in early June 2006, and follow-up reminder emails were sent again in August and October 2006. Participants clicked on a hyperlink in the email message to access the survey web site, and they completed an informed consent agreement before answering survey items. The survey required approximately 30 minutes to complete.

Response rates: Of the original 745 individuals who received an invitation to participate, a total of 49 had email addresses with permanent communication failures. For the remaining 696, it was assumed that the invitation email was delivered successfully. Of these 696, 202 individuals (29%) logged on to the survey web site, while 176 (25%) entered and completed the survey. Response rates varied slightly from one section of the survey to the next, with 157 (23%) being the lowest number of responses for any one section.
IV. Survey Results
1. Demographics of survey respondents

(a) Institution

Institutional setting

Public 48%  Private 49%  Other 2%

Institutional size

Less than 1000 11%  1000-4999 32%  5000-9999 2%  15,000-19,999 14%  20,000 or more 11%

Institutional type

Grad & 4-year 31%  4-year degree 1%  2-year degree 67%

(b) Program

Language program locus

Department 7%  Specific section 4%  College or school 6%

Language program composition

Conglomerate-languages program 15%  Single-language program 16%  Language-family program 69%

Program subcomponents

Language certificate 67%  PhD 1%  Heritage language 4%  Teaching certificate 2%  Interdisciplinary degree 31%

Languages taught: Respondents identified a total of 94 different languages on offer in their programs. Of these, 39 languages were not offered in more than one program. The 10 languages that were identified most frequently are shown in the following table.

Language program enrollments per semester

<table>
<thead>
<tr>
<th>Language program</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>846</td>
<td>918</td>
<td>18</td>
<td>5000</td>
</tr>
<tr>
<td>Spanish</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>German</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Italian</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Russian</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Japanese</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Chinese</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Portuguese</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Arabic</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Latin</td>
<td>8</td>
<td>3</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Respondents' role in the program

Chair/head of dept 83%  Coordinator of prog component 8%  Director of multiple lang prog 8%  Head of language section 3%  Professor/Instructor 3%

The AVERAGE respondent to this questionnaire was (a) a chair of a foreign language department that consisted of multiple language sections (in particular: French, Spanish, German); (b) offered majors, minors, summer courses, language requirements, and study abroad, with a total of around 900 enrollments per semester; and (c) granted undergraduate and graduate degrees in either a private or a public institution of medium size.
2. Impetuses for program evaluation

Question: “To what extent do you feel under pressure from the following SOURCES to engage in program evaluation?”

A total of 63 respondents provided additional comments regarding impetuses for evaluation. The 3 most frequent comments were in these categories:

1. Accreditation/administration demands (N = 11), for example:
   “As part of its re-accreditation, the University has required all undergraduate programs to create and implement outcomes-oriented assessment plans.”

2. Individual/group professional ethic to evaluate (N = 10), for example:
   “We have a social and moral responsibility towards our students and towards society at large to state as clearly as we can what it is that we do for them and why what we do is valuable.”

3. (tie) Trend in higher education (N = 4), for example:
   “General trends in education, esp. focus on learning outcomes & student centered instruction.”

3. (tie) Basis for hiring positions (N = 4), for example:
   “All departments are required to undergo a review in order to request a new tenure-track position.”

Other impetuses receiving lower ratings (ranging from M = 2.52 to 1.28):
tenure and promotion, individual faculty, teacher accreditation, professional organization, students, employers, external funder, student government, parents, and community.

3. Program elements under focus in evaluation

Question: “For the following program COMPONENTS, please indicate the extent to which each (a) is CURRENTLY evaluated in your FL program, and (b) SHOULD BE evaluated in your FL program.”

A total of 55 respondents provided additional comments regarding program elements under focus in evaluation. The 3 most frequent comments were in the following categories:

1. Study abroad (N = 9), for example:
   “Overseas programs, their quality, their coordination with the on-campus program.”

2. Role or value of the program (N = 6), for example:
   “The program's role in liberal arts education, which is not quite the same as our mission/goals because it comes from outside the department.”

3. Faculty performance (N = 4), for example:
   “We already also evaluate scholarly and research efforts by FL faculty.”
4. Uses for evaluation

Question 1: “For the following evaluation APPLICATIONS, please indicate the extent to which evaluation (a) is currently used, and (b) should be used.”

How program evaluation is currently versus should be used

- 1. Understanding/improving how your FL program is functioning
- 2. Understanding/improving the outcomes of your FL program
- 3. Understanding/improving the value/worth of your FL program
- 4. Justifying/defending your FL program
- 5. Getting your FL program accredited
- 6. Raising awareness about your FL program
- 7. Holding your FL program accountable
- 8. Comparing your FL program with other programs
- 9. Informing personnel decisions within your FL program
- 10. Improving FL education on the whole

Overall evaluation use

Overall current capacity (M = 2.94, SD = 0.77)

A total of 20 respondents provided additional comments regarding the uses for program evaluation. The 3 most frequent comments were in these categories:

1. Improving curriculum (N = 5), for example:
   “To assess the undergraduate curriculum and improve it by understanding what your students want.”

2. Hiring faculty (N = 3), for example:
   “Decisions about hiring -- whether we will get a new line or not if someone retires.”

3. Improving faculty performance (N = 2), for example:
   “It is hard in small department to give faculty evaluations of their teaching/collegeship/research--we're all tenured. It needs to come from outside the department. That might improve our collective performance.”

5. Program evaluation methods and capacities

Question: “How useful are the following METHODS, TOOLS, and TECHNIQUES for meeting evaluation needs in your FL program?”

Top 5 most useful program evaluation methods

Other methods receiving lower ratings (ranging from M = 3.03 to 2.00): proficiency assessments, course evaluations, portfolio assessments, grades, longitudinal tracking, interviews, questionnaires, self-assessments, journals, and focus groups

A total of 30 respondents provided additional comments regarding the methods of program evaluation. The 3 most frequent comments were in the following categories:

1. Standardized/national exams (N = 5), for example:
   “A standard national test will come in handy.”

2. External reviews (N = 3), for example:
   “External evaluations by competent faculty from other high quality language programs.”

3. Exit/comprehensive exams (N = 3), for example:
   “Oral and written senior (major) comprehensive examinations.”
**6. Helpful program evaluation resources**

*Question: “Please indicate which of the following RESOURCES would help you to conduct useful evaluations.”*

<table>
<thead>
<tr>
<th>Resource Description</th>
<th>Percentage</th>
<th>Mean (M)</th>
<th>Standard Deviation (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Concrete examples of FL program evaluation (M = 3.50; SD = 0.73)</td>
<td>60%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>9. Program-internal expertise in FL program evaluation (M = 3.27; SD = 0.82)</td>
<td>46%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>15. Systematic approaches for putting evaluation data to use (M = 3.22; SD = 0.91)</td>
<td>47%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>4. Templates for FL program evaluation process (M = 3.22; SD = 0.86)</td>
<td>44%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>1. Professional development workshops (M = 3.19; SD = 0.89)</td>
<td>45%</td>
<td>5%</td>
<td></td>
</tr>
</tbody>
</table>

**Other resources receiving lower ratings** (ranging from $M = 3.18$ to $2.41$):
- Data-collection tools
- Assessment workshops
- Awareness-raising strategies
- Professional organization support
- External consultant
- Data analysis software
- National evaluation resource center
- Online help service
- Textbooks
- Evaluation coursework

**A total of 21 respondents provided additional comments regarding resources helpful for conducting evaluations. The most frequent comments were in the following category:**

1. **Time/money/staff** to do evaluation ($N = 7$), for example:
   
   “Concrete ideas that are time-efficient and use what faculty are already doing and department goals as basis.”

2. **Other comments reflected the program-specific nature of resource needs, as in the following example:**
   
   “In Slavic, we would benefit by our professional organization (AATSEEL) discussing and working towards a statement of desired and reasonable outcomes on multiple parameters based on degree types. Perhaps other languages have already worked this out, but it has been difficult to engage our colleagues in such a discussion. Any strategies welcome.”

**7. Concerns with program evaluation**

*Question: “What main concerns do you have, if any, about evaluation in your FL program?”*

A total of 113 respondents provided comments regarding their concerns with evaluation. The most frequent comments were in the following categories:

1. **Lack of time, overburdened faculty** ($N = 26$), for example:
   
   “Limited personnel and time resources versus the personnel and time needed to carry out assessment and program evaluation. Our university lacks a distinct office of institutional research to fully support assessment and data analysis.”

2. **Use, usefulness, follow-through on evaluation** ($N = 17$), for example:
   
   “Evaluations results should be used for concrete results and not just because it is done elsewhere. They should also be specifically tailored to each department.”

3. **Willingness of faculty to participate** ($N = 16$), for example:
   
   “Lack of interest in evaluation of their programs by most faculty. Many are hostile even to the idea of needs analysis.”

4. **Institutional support, funding, help** ($N = 16$), for example:
   
   “Until the administration, often the 'champion' of evaluation and accountability, provides the resources necessary to make evaluation more than just a quick exercise to meet minimal benchmarks, it will not improve.”

5. **Understanding, knowledge, expertise** ($N = 10$), for example:
   
   “One of the key issues is developing a broad understanding of evaluation and assessment issues in the faculty body as a whole so that there is a continuous attention to these issues.”

6. **Fear, misuse by external forces** ($N = 9$), for example:
   
   “Evaluation from administrators or colleagues from other departments on campus has often demonstrated how little they know about FL teaching today.”

7. **Available instruments and procedures** ($N = 8$), for example:
   
   “We currently do not have any good assessment mechanisms to determine whether our seniors (majors & minors) are adequately prepared in their respective languages.”

8. **Comparability of evaluation data** ($N = 8$), for example:
   
   “Evaluation is not comprehensive (some language areas receive none). Data is not always comparable between language areas. Methods and rubrics of data gathering are not consistent.”
8. Executive summary

Impetus to evaluate: relatively high demands to evaluate

- Top-down impetus (e.g., university administration, dean, and accreditation) and sense of professional responsibility (e.g., department faculty as a whole, individual ethic) are the two major driving forces for program evaluation. These potentially compete in determining the forms and uses of evaluation.
- Not much pressure from stakeholders (e.g., community, parents, students) other than educational community and program administration.

Program evaluation focus: somewhat broader program coverage is needed

- More attention needed: Three program elements (teacher development, students’ learning needs, students’ attitudinal change) are least evaluated & call for more.
- Already happening: Programs already have teacher evaluation in place (highest priority, least increase), utilizing classroom observation and/or course evaluation.
- Important focus: Curriculum (e.g., mission, goals, curriculum scope & sequence) and student learning outcomes are also high priorities for evaluation focus.

Program evaluation use: substantial increase in use is sought

- More understanding and improvement: Through program evaluation, respondents want not only to improve FL program functions, outcomes, and value, but also the field of FL education as a whole.
- Less justifying and accrediting: Program accreditation is the least desired use, and justifying/defending the program is perceived to be most over-used, indicating considerable dissonance between mandated evaluation practice versus the perceived value of evaluation for program improvement.

Program evaluation methods and capacity: useful methods, diverse capacities

- Most useful tools: Program-specific, internal practices, like classroom observations, outcomes assessment, meetings, document analysis, and exams.
- Least useful tools: Tools that are more time consuming, less familiar, and/or hard to analyze/interpret: focus groups, journals, and self-assessment. Course evaluations are often conducted via questionnaires, however, 40% of the respondents perceived these as not very useful.
- Program evaluation capacity: 28% of the respondents indicated no capacity or little capacity; 24% indicated a lot of capacity.

Useful resources: targeting efficient, use-oriented strategies

- Most helpful: More concrete examples, templates, and models of systematic use-oriented evaluation are sought.
- Mixed reviews: Face-to-face and less time-consuming services are welcomed (e.g., professional development workshops) but not coursework or textbooks on program evaluation that require potentially greater time commitment.
- Critical capacity development need: Most programs value internal faculty expertise in program evaluation as a useful resource, coupled with time and institutional support to engage in evaluation.

Implications

- Professional ethic and responsibility towards program evaluation suggest awareness of its potential value for understanding and improving FL education.
- Tensions between internal versus external impetuses for evaluation call for careful attention to the educational impact of resulting practices.
- Across the U.S., college FL programs have diverse evaluation capacities, concerns, and needs, though most would like to see increased use of evaluation but most face time and resource constraints.
- There is a clear need for strategies to leverage institutional awareness and support, as well as faculty engagement, if evaluation is to play a legitimately useful role in FL programs.
- Evaluation methods should prioritize efficiency, systematicity, and local utility, though there is some desire for standardized evaluation instruments.
- More concrete cases, examples, tools, and use-oriented models that are specific to various types of language programs (distinct languages, components, institutional settings) are prioritized for dissemination.
- Short-term evaluation capacity development might best be addressed through program-specific workshops, direct assistance/outreach, and dissemination of good examples; Long-term capacity development likely depends on awareness raising, knowledge generation, and discipline-wide attention to professional development in evaluation.
V. Foreign Language Program Evaluation Resources

1. Currently available [http://www.nflrc.hawaii.edu/evaluation/]

- Thematic & searchable annotated bibliography on language program evaluation
  - Language program evaluation books, chapters, and articles are summarized and categorized into a searchable data-base, featuring key themes, methods, and examples of evaluation.

- Using surveys in language program evaluation
  - Recommended readings and references of surveys used in foreign and second language program evaluations.
  - Examples of surveys used in foreign and second language program evaluations.
  - A list of references on survey methodology books, evaluation studies that used surveys, and online survey research.
  - A list of websites that introduce survey design, data collection methods, and data analysis.
  - Comparison of available online survey tools.

- Using interviews and focus groups in language program evaluation
  - Recommended readings and references on interviewing, focus group techniques, and qualitative data analysis.
  - A list of websites on designing and analyzing interviews, advice on focus group techniques.
  - Examples of language program evaluations that employed interviews and focus groups.

2. Coming soon….

- Templates
  - A step-by-step guide to prioritizing uses for evaluation in college FL programs
  - Student learning outcomes assessment: A use-driven design template

- Models
  - Successful models for staffing and funding assessment and evaluation efforts in U.S. colleges
  - Realistic timelines for language program evaluation: Less is more
  - Increasing participation in (and the value of) evaluation: Models of practice

- Examples
  - Collected examples of teacher development evaluation in college FL programs
  - Collected examples of learner needs analysis in college FL programs
  - Collected examples of outcomes assessment in college FL programs

VI. Opportunities to participate and learn more!

- May 28 - June 6, 2007: Summer Institute at the University of Hawaii National Foreign Language Resource Center. "Developing useful evaluation practices in college foreign language programs." Join colleagues and evaluation experts for a two-week intensive workshop on language program evaluation. For more information, see: [http://www.nflrc.hawaii.edu]

- June 6 - June 10, 2007: Summer Seminar West of the Association of Departments of Foreign Languages, at the University of Hawaii. The annual meeting of the ADFL will feature a focus-day on language program evaluation, with expert panels, examples of good practice, and evaluation-related working sessions. For more information, see: [http://www.adfl.org]

- Ongoing: Professional education in language program evaluation. The University of Hawaii SLS Department offers a Master’s degree with specialization in “Language Assessment, Measurement, and Program Evaluation” (see [http://www.hawaii.edu/sls/lampe]), as well as a one-year Advanced Graduate Certificate program for professionals who want to receive individualized training in related topics.

Please monitor the Foreign Language Program Evaluation Project web site ([http://www.nflrc.hawaii.edu/evaluation]) for information on future presentations, workshops, and other opportunities to participate in and/or contribute to this project. If you would like to propose an event or make any other suggestion related to FL program evaluation, we would love to hear from you. Send us an email ([jnorris@hawaii.edu, yukikow@hawaii.edu]).
FOREIGN LANGUAGE PROGRAM EVALUATION PROJECT
University of Hawai‘i at Mānoa
Department of Second Language Studies
1890 East-West Rd.
Honolulu, HI 96822

Please visit our project website:
http://www.nflrc.hawaii.edu/evaluation

jnorris@hawaii.edu

Copyright © 2005 John M. Norris & Yukiko Watanabe

Mahalo!