1. Background

1.1. Clarification of the terms outcomes-based evaluation and outcomes assessment

• **Assessment:** “the systematic gathering of information about student learning in support of teaching and learning”

• **Evaluation:** “the gathering of information about any of the variety of elements that constitute educational programs, for a variety of purposes that primarily include understanding, demonstrating, improving, and judging program value…brings evidence to bear on the problems of programs, but the nature of that evidence is not restricted to one particular methodology” (Norris, 2006, p. 579)

→ Student learning outcomes assessment is basically outcomes-based program evaluation focused on student learning. Evaluation > outcomes assessment (the purpose is to effect program-level decisions)

→ In higher education, the term “assessment” is often used instead of program evaluation to include evaluation of many program elements (beyond student learning). → causing confusion

1.2. What are outcomes?

• The effect of program activities and deliverables on target audiences of a program.

• Often described in terms of change in knowledge, skills, attitudes, beliefs (→ learning outcomes), program/social policies and conditions.

1.3. Why outcomes?

• **Accountability and accreditation**

  Many language programs in higher education are now concerned with issues of accreditation, accountability to funders, or both. In both accreditation-driven and accountability-driven evaluation, the focus has turned to outcomes. This is because funders and accrediting organizations are no longer satisfied with programs reporting on inputs (resources that have been put into a program), or outputs (activities the program conducted), they want to know about results. What effect did the program have?

• **Positive impact on programs**

  Table 1 (next page) provides examples of the many positive impacts outcomes-based evaluation can have on a department or program beyond simply meeting accreditation or accountability demands.
Table 1. *Example Evaluation Impact Types from the Outcomes Assessment and Evaluation Literature*

<table>
<thead>
<tr>
<th>Focus</th>
<th>Types</th>
</tr>
</thead>
</table>
| **Program** | • Program growth  
- Increase in enrollment and greater number of courses (Milleret & Silvera, 2009)  
• Change in project focus and objectives  
- Initial needs analysis redefined the aim of the project (Coleman, 1992)  
• Change in curriculum  
- Realignment of the curriculum to conform to prog-external standards (Carstens-Wickham, 2008)  
- Improved curricular articulation (Byrnes, 2002)  
- Standardized outcomes (Milleret & Silvera, 2009)  
- Improved strategies for providing extra support (Elder, 2009)  
• Change in teaching and learning  
- Assessment integration into teaching and learning (Byrnes, 2002)  
- Greater student satisfaction with their classroom experience (Milleret & Silvera, 2009)  
• Change in assessment practices  
- Standardized classroom assessment practices (Milleret & Silvera, 2009)  
- Changes in program selection or placement practices (Elder, 2009) |
| **Organization** | • Better organizational vision  
- Change in strategic planning, i.e., new and re-focused priorities (Carstens-Wickham, 2008)  
• Better collaboration  
- Enhanced can-do attitude and positive feeling toward team work (Gau Sempere, et al., 2009)  
• Increase in resources and support  
- Increase in salary and support for professional development (Bernhardt, 2006)  
• New understanding and awareness about the program  
- New awareness of the importance of the National Standards (Carstens-Wickham, 2008)  
- Better understanding about students (Milleret & Silvera, 2009)  
- Better understanding about departmental identity (Gau Sempere, et al., 2009)  
- Better understanding of the contextual and political factors influencing a program (Harris, 2009)  
- Realization of the need to have similar goals, standards, and practices across different languages to create a cohesive department. (Gau Sempere, et al., 2009)  
• Sense of ownership of the program and empowerment  
- TAs (and students) felt their voice were heard and grew a feeling of ownership of the program (Milleret & Silvera, 2009)  
• Evaluation capacity building: More positive attitudes and perceptions towards evaluation, increased knowledge about evaluation  
- Increased faculty interest in evaluation (Carstens-Wickham, 2008)  
- TAs obtained positive attitudes about evaluation and learned tools to conduct evaluation (Milleret & Silvera, 2009)  
- Teachers involved in evaluation gained evaluation skills (How to plan evaluation, collect and interpret data, etc.) (Alderson & Scott, 1992) |
| **External organization** (or Inter-organizational relationships) | • Increased communication and collaboration between organizations  
- Increased vertical (upper admin) collaboration (Walther, 2009)  
- Dynamic interaction between program units influenced language attitudes and behaviors of other units (Coleman, 1992)  
• Increased visibility of the department in the institution  
- Increased visibility of the department on campus as a leading department on assessment (Grau Sempere, et al., 2009) |
2. Useful Outcomes-Based Evaluation in FL Education - Gov’t Grant Funded Contexts

2.1 The call for useful outcomes-based evaluation in government grant funded programs

- **British Council / British Overseas Development Administration projects**
  Much of the early language program evaluation literature grew out of work on outcomes-based evaluations of large-scale language teaching projects funded by the British Council in the 1980’s and 90’s. The evaluators of these projects, disillusioned by the ineffectiveness of their evaluations and reflecting on lessons learned from their experiences, called for more focus on use and usefulness in evaluation.
  - Beretta - (CLT, Bangalore, India)
    Reflecting on the Bangalore Evaluation stated:
    *If I had known then what I know now...I would have oriented the initial discussion to a consideration of what use could be made of the findings....I would have... pursued the pre-evaluation discussion in order to arrive at a clearly articulated agreement about the use that was to be made of the findings (and thus what kinds of information would have leverage and whether or not such data could be collected)...* (Beretta, 1992, p. 264)
  - Mackay - (ELTP, Indonesia)
    Argued that an evaluation is only worth doing if the principal stakeholders can show that the evaluation “can be put to profitable use” and “is likely to be put to profitable use” (Mackay, 1988, p. 40).
  - Alderson & Scott – (Brazilian ESP Project)
    In his Guidelines for the Evaluation of Language Education (1992), Alderson pointed out:
    *Evaluations are perhaps best distinguished from ‘research’ in that they are intended to serve practical ends, to inform decision makers as to appropriate courses of action, and, above all, to be useful and to be used. An evaluation that is not used is in some important sense a failure* (p. 298).

- **Review of the Title VI international centers**
  - More recently, the National Research Council’s 2007 review of Title VI projects also emphasized the need to focus on both outcomes and use:
    *Meaningful evaluations of outcomes and impacts are lacking. More rigorous, periodic evaluations are needed to ensure public accountability, inform the process of regular reauthorization, and advance continued departmental efforts to stimulate internationalization* (p. 8).
    *Recommendation: The Department of Education should commission independent outcome and impact evaluations of all programs every 4 to 5 years* (p. 9).
  - The report also remarked that the data collection system used by the Dept. of Ed was viewed as a “burdensome, time consuming requirement, not a resource,” and that there was a lack of information about the “purpose and use of required data.” Furthermore, it stated that in the development of the new data-collection system “consideration was not given to whether the data collected should be collected and how it would be used for program monitoring or improvement purposes” (p. 218).

- **2010 Title VI grant applications**
  The 2010 grant applications for National Resource Centers, Language Resource Centers, and Foreign Language and Area Studies Fellowships all placed new emphasis on program evaluation, and particularly outcomes-oriented evaluation
  - Dear Applicant Letters for LRCs and NRCs emphasized quality evaluation plans that included outcomes:
    *Finally, I would like to share with you the importance of ensuring that your application includes a strong evaluation plan*...
    *The evaluation should also, where appropriate, attempt to link program outcomes to specific activities or materials.*
Evaluation sections of the LRC, NRC, and FLAS applications were worth approximately 20% of the application points

Criteria 9C of the technical review form for NRCs (10% of application points)

Does the applicant provide an evaluation plan that is comprehensive and objective and that will produce quantifiable, outcome-measure-oriented data?

Both NRCs and LRCs required to include Appendix 4

Provide a table of the objectives that will guide project activities for the FY 2010-2013 performance period...Objectives should define the outcome (effect or benefit) to be realized.

As educators, administrators and evaluators, how do we meet the demands for outcomes-based evaluation, while at the same time learning from the lessons of earlier evaluations and focusing on use and usefulness?

2.2. Outcomes at work: NRC evaluation project sample

- **Project**
  A new, two-year language program for an LCTL
  - A collaborative effort between a UH language department and an NRC. The NRC provided the funding to seed the program.

- **Evaluation needs and impetus**
  **External:**
  - New grant requirements for NRC 2010-2014 grant cycle
  **Internal:**
  - A desire by the NRC to understand the language department’s goals for the new language program.
  - Concern that the university did not pick up funding for the language program at the end of the previous grant cycle
  - Interest on the part of the department chair to develop SLOs for all language programs in the department

- **Type of outcomes**
  **Program-level outcomes**
  - Impact on the student population, university, and heritage language community
  **Course-level outcomes**
  - SLOs / Language proficiency

- **Who’s involved?**
  - Evaluator
  - Language department faculty: Chair and language instructor
  - NRC Faculty: Outreach coordinator

- **Evaluation steps and activities**
  - Identifying evaluation needs - Meetings
  - Stating and mapping project objectives and outcomes - Evaluation table
  - Developing an evaluation plan - Prioritize, develop timeline
  - Developing SLOs - “Talk Story”
  - Assessing SLOs - Survey
  - Gathering data on student backgrounds and interests – Survey
  - Interpreting survey results - Meetings
  - Developing an action plan – Meetings, recorded in the evaluation report
  - Reporting – Member checking, final report submitted to NRC

- **Impact strategies**
  Impact on programs and projects
  - Explicit separation of skill levels and expectations
Impact on organization
  - Clarification / Understanding of program goals and purpose
  - Improved communication between collaborators
  - Increased sense of involvement and collaboration
  - Identification of resource gaps
  - Improved planning for similar program in initial stages
  - Improved communication with new instructor for similar program
  - Increased interest in evaluation

Impact on the relationship with external programs and organizations
  - Ability to articulate the purpose and value of the program to external stakeholders
  - Able to meet external evaluation demands

- Reflections on particular actions or techniques that enhanced impact and transformation
  - Focus on feasibility - Long-term planning, limited scope in the first year, respect for faculty workloads
    = Sustained stakeholder involvement and motivation
  - Iterative process - Allowed for drafting and re-drafting of objectives, outcomes, SLOs and evaluation plan as we learned more about the program
    = Non-threatening, allowed for ongoing improvement, learning, and adjustments
  - Support from leadership – NRC program staff and department chair strongly supportive of evaluation work, willing to make time in schedule and plan for evaluation
    = Important decision-makers and stakeholder involved, instructors motivated to participate, sufficient time for planning and implementation
  - Stakeholder involvement – Chair and instructor heavily involved
    = Improved understanding of the program, evaluative know-how, buy-in to the evaluation, evaluation use and impact

- Reflections on particular actions or techniques that limited impact and transformation
  - Part-time, adjunct instructor, not paid for participation in evaluation activities
    = Limited the scope and pace of the evaluation
  - Lack of clarity about participant roles
    = Some feelings of uncertainty, participants wondering if they should do more

3. College FL Programs: Accreditation-Driven Outcomes Assessment

3.1. Regional accreditors’ focus on examining student learning outcomes
  Mid-80s: Government’s interest on return on investment (e.g., A Nation at Risk, NCEE, 1983)
  → DOE’s call for “Educational Effectiveness” review
  → By early 1990s: Accreditation included standards on educational effectiveness

3.2. National survey on program evaluation in tertiary foreign language programs (Watanabe, Norris, González-Lloret, 2009)
  - Top 2 purposes: understanding/improving how the program is functioning; understanding/improving the outcomes of the program
  - Issues & concerns:
    - Evaluation use: Misuse and abuse of evaluation findings, lack of utility
    - Evaluation capacity: Limited time & resources, resistance & buy-in

3.3. Call for useful outcomes assessment in higher education
  - Norris (2000, 2005) and Banta (2002) suggests Patton’s utilization-focused evaluation model as an approach to engage people in outcomes assessment, so that programs will use the assessment findings for intended purposes.
3.4. Useful outcomes assessment at work (two examples)

- **Project context**: A 2-year student learning outcomes (SLO) assessment project involving 8 FL programs (multi-language or a single-language department/unit) across two universities.

- **Project goals**
  - Research goal: To understand facilitating and hindering factors on planning, design, implementation, and use of outcomes assessment in college foreign language programs.
  - Local project goals: To produce locally useful outcomes assessment practices; to enhance assessment/evaluation capacity within the program; and to create a sustainable outcomes assessment system.

- **Case studies**: High assessment capacity program and limited assessment capacity program (Table 2)

Table 2. Outcomes Assessment Context, Activities, Impact, and Strategies

<table>
<thead>
<tr>
<th></th>
<th>HIGH assessment capacity program</th>
<th>LIMITED assessment capacity program</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational features</td>
<td>• Multi-language (4 languages), large department</td>
<td>• Multi-language (3 languages), small department</td>
</tr>
<tr>
<td></td>
<td>• Curriculum coordinators</td>
<td>• Decision-making: Unsteady in the past</td>
</tr>
<tr>
<td></td>
<td>• Chair: Strong leadership, designates power to committee, advocates assessment</td>
<td>• No coordinators, no curricular discussion/coordination</td>
</tr>
<tr>
<td>Assessment history</td>
<td>• Experience going through one cycle of assessment, focusing on oral proficiency.</td>
<td>• Chair turnover; new chair (1/2 Dean’s office, 1/2 chair)</td>
</tr>
<tr>
<td></td>
<td>• Assessment procedures established.</td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td>- Perceived high assessment capacity</td>
<td>- Perceived low assessment capacity</td>
</tr>
<tr>
<td></td>
<td>- Willingness to engage in assessment work</td>
<td>- Willingness to engage in assessment work</td>
</tr>
<tr>
<td><strong>Working Group</strong></td>
<td>A few TT + NTT faculty, and subcommittees</td>
<td>All faculty (TT+NTT) as a WG. Lang center director (external to the program) became the assessment coordinator.</td>
</tr>
<tr>
<td>Assessment focus</td>
<td>Writing outcomes for the lang. requirement program</td>
<td>BA major program outcomes</td>
</tr>
<tr>
<td>Assessment use:</td>
<td>(1) To showcase and standardize writing pedagogy &amp; assessment</td>
<td>Assessment focus:</td>
</tr>
<tr>
<td></td>
<td>(2) To adjust and showcase language requirement writing outcomes</td>
<td>BA major program outcomes</td>
</tr>
<tr>
<td>Activities</td>
<td>• Gather existing prompts, rubrics, and writing samples.</td>
<td>Assessment use:</td>
</tr>
<tr>
<td></td>
<td>• Specify learner performance profile for the 200-level courses.</td>
<td>• Validity testing: Whether target proficiency levels of the newly created outcomes statements are appropriate</td>
</tr>
<tr>
<td></td>
<td>• Create/adjust a writing task that captures the outcomes.</td>
<td>• Showcasing of assessment work to external program reviewers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Began with situating assessment work (capacity building, getting everyone on the same page, airing out concerns, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Stating outcomes: Individual + peer institutions --&gt; consolidated + consensus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data gathering: ACTFL OPI + WPT</td>
</tr>
</tbody>
</table>
### Impact, factors, and strategies

<table>
<thead>
<tr>
<th>Formed an inclusive working group</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Increased communication between TT and NTT faculty</td>
</tr>
<tr>
<td>→ Allowed a structural change: democratic decision making (involvement of NTT)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Going through one SLO assessment cycle (Assessment use and program impact were observed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Increased confidence in and better attitude toward SLO assessment (willingness to engage in evaluation)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategic focus on writing to cultivate interests</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Cultivated interests from TT faculty</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Provide readings and resources to build assessment capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Increased confidence and knowledge in SLO assessment. Able to explain assessment framework and approaches to other programs.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Informing upper administration about assessment work</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Monetary support for assessment capacity building and assessment effort</td>
</tr>
<tr>
<td>→ Elevation of “assessment profile” of facilitators/individuals/organization</td>
</tr>
<tr>
<td>→ Individual faculty members were invited to participate in institutional level assessment effort. Upper administration plan to showcase programs engaged in assessment to the accreditors.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entire faculty participation and opportunities for curricular discussion and sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Generated programmatic thinking</td>
</tr>
<tr>
<td>→ Shared understandings of program outcomes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Air out concerns &amp; perceptions about assessment. Emphasis on bottom-up approach and internal use</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Untangled misunderstandings and gained buy-in from faculty to be involved.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of quick data gathering and credible evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Realization of the usefulness of external standardized assessment data to understand the program.</td>
</tr>
<tr>
<td>→ Found variability in levels for one of the program, due to heritage background</td>
</tr>
<tr>
<td>→ Assured appropriate target language proficiency level for the majors.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lessons learned from data misuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Climate: Program review pressure and excitement to obtain proficiency data</td>
</tr>
<tr>
<td>→ Ethical consideration: Data was reported with course information and instructor names.</td>
</tr>
<tr>
<td>→ Utility consideration: The Dean was not informed of the program’s intended assessment use, and received the raw data reported. The Dean misused the data to explain teacher effectiveness to the external reviewers. The Dean and the coordinator learned that reporting requires careful considerations of intended use and ethics.</td>
</tr>
</tbody>
</table>

### 4. Strategies for Enhancing Impact

- **Organizational Leaders /Evaluation Facilitators (Dept chair, Center director, etc.)**
  - Recognize the external and internal evaluation needs.
  - Involve stakeholders in evaluation decision-making.
  - Create a safe environment for participants to ask questions and voice concerns, encourage faculty collaboration.
  - Be realistic about scope, resources, timelines, etc.
  - Seek support and opportunities for developing faculty and staff for evaluation training.
  - Recognize that evaluation is essential for organizational decision making about their program.
  - Recognize faculty effort and find ways to showcase their work at conferences and internal professional development events.
  - Communicate with and inform upper administration about the evaluation work and necessary resources to launch, implement, and sustain evaluation.
  - Clarify roles and responsibilities, assign one person who can closely monitor evaluation implementation. Maintain consistent meeting times for evaluation.
  - “Walk the walk”

- **Committee members**
  - Prioritize evaluation focus based on intellectual and program needs.
  - Make use of faculty research expertise.
  - Consider ways to ensure credible evidence. (Tap into expert knowledge and resources for instrumentation, data gathering, analysis, and interpretation.)
  - Participate and provide feedback wherever possible.
  - Voice concerns, questions, and needs.
  - Respect others’ concerns, questions, and needs.
  - Maintain focus on use at all stages of the evaluation.
5. References


Harris, J. (2009). Late-stage refocusing of Irish-language programme evaluation: Maximizing the potential for productive debate and remediation. Language Teaching Research, 13(1), 55-76.


