Brigham Young University

BYU ScholarsArchive

2005-12-02

A Metaevaluation of An Evaluation of A Second Language Course

Edmilson Bezerra Torres
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Educational Psychology Commons

A METAEVALUATION OF AN EVALUATION

OF A SECOND LANGUAGE COURSE

by

Edmilson B. Torres

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APROVAL

Of a project submitted by

Edmilson B. Torres

This dissertation has been read by each member of the following graduate
committee and by majority vote has been found to be satisfactory.

_____   _____
Date   David D. Williams Chair

_____   _____
Date   Andrew S. Gibbons

_____   _____
Date   Paul F. Merrill

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Edmilson B. Torres in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____   _____
Date   David D. Williams
                                                                       Chair, Graduate Committee

Accepted for the Department   _____
                                                                       Andrew S. Gibbons
                                                                       Department Chair

Accepted for the College   _____
                                                                       Richard Young
                                                                       Dean, David O. McKay School of Education

ABSTRACT

A METAEVALUATION OF AN EVALUATION

OF A SECOND LANGUAGE COURSE

Edmilson B. Torres

Department of Instructional Psychology and Technology

Master of Science

This project is a metaevaluation or critique of an evaluation of an intermediate course in a second language curriculum. In this report, the intermediate course evaluation is described to provide a basis for understanding the metaevaluation. Then the metaevaluation is presented.

The evaluation was the first stage of a department-approved and college-supported curriculum redesign project to improve the quality of a second language curriculum in terms of instructional materials, methodological approaches, and pedagogical practices to promote optimal second language learning gains. Through the evaluation, strengths and weaknesses of the curriculum were identified so that these issues could be addressed during curriculum development.

This metaevaluation identifies strengths and weaknesses of the intermediate course evaluation in terms of its utility, feasibility, propriety, and accuracy.

Recommendations are made for improving the evaluation. In addition to several strengths, the metaevaluation identified some ways the intermediate course evaluation could be improved. Findings include the need for creating better relationships with stakeholders by more clearly understanding their concerns and views, creating more comprehensive contracts, and clarifying responsibilities and rights pertaining to the use of reports and data.

ACKNOWLEDGEMENTS

Table of Contents

List of Appendices

Introduction

This is a report of a metaevaluation or an evaluation of an evaluation. Metaevaluations are conducted by metaevaluators who observe and critique primary evaluations' procedures against sets of standards. In this study, the metaevaluation was conducted of an evaluation of an intermediate language course to address concerns of the metaevaluator's masters degree committee, as well as standards for program evaluation promoted by professional evaluators, including utility, feasibility, propriety, and accuracy.

To set the context for metaevaluating with these standards, a brief literature review summarizes key issues and then the intermediate course evaluation, which was the object of the metaevaluation, is summarized. The intermediate course evaluation summary identifies the evaluand or object being evaluated as an intermediate language course. Further, it clarifies background information about the course and its evaluation, who the stakeholders were, their criteria for judging the course, the questions answered, data collection and analysis procedures, reporting strategies, results, recommendations, schedule, and budget.

With this context established, the metaevaluation of the intermediate course evaluation is presented in terms of how well it met thirty criteria associated with the utility, feasibility, propriety, and accuracy standards. In particular, through observations, report reviews, and interviews, the metaevaluator gathered data regarding the completion of the intermediate course evaluation and judged it in light of these standards.

The utility standards include seven criteria: stakeholder identification, evaluator credibility, information scope and selection, values identification, report clarity, report

timeliness and dissemination, and evaluation impact. The feasibility standards include three criteria: practical procedures, political viability, and cost effectiveness. The propriety standards include eight criteria: service orientation, formal agreements, rights of human subjects, human interactions, complete and fair assessment, disclosure of findings, conflict of interest, fiscal responsibility. Finally, the accuracy standards include twelve criteria: program documentation, context analysis, described purposes and procedures, defensible information sources, valid information, reliable information, systematic information, analysis of quantitative information, analysis of qualitative information, justified conclusions, impartial reporting, and metaevaluation. Based on these thirty criteria, the metaevaluation identifies strengths and weaknesses of the intermediate course evaluation and offers recommendations for improving future evaluations.

Organizationally, this project begins with a literature review, followed by a brief explanation of methods used to conduct the metaevaluation, a description of the primary evaluation or the intermediate course evaluation which was metaevaluated, and then the metaevaluation itself.

Literature Review

Evaluation is a well established process for helping teachers and educational programs improve courses and curricula. Scriven (1981) expressed the concern of many about the importance of evaluation supporting course and curriculum design when he noted,

> If we wanted to know about the quality of the curriculum, we would need to examine the content that was being taught. What are the central ideas constituting the curriculum? What concepts are focused upon? What general theoretical structures are being offered to the students on which these concepts can be placed. (p. 42)

Lynch (1996) implies that program evaluation should play an essential role in the development of applied linguistics, such as the intermediate course which was the evaluand in this study. He clarifies that the evaluation process should be focused around stakeholders and their values, "Identification of the evaluation audience leads to determining the evaluation goals, or purpose" (p. 3). Lynch (2003) further reinforces this view by noting that stakeholders are "the audiences that should legitimately have a voice in determining the goals for the assessment and evaluation" (p. 16).

The discrepancy evaluation model proposed by Provus (cited by Brown, 1989) fit the need of the intermediate course evaluation well. As Provus noted:

> Program evaluation is the process of 1) defining program standards, 2) determining whether a discrepancy exists between some aspect of program performance and the standards governing that aspect of the program, and 3) using

discrepancy information either to change performance or to change program

standards. (p. 228)

Furthermore paying attention to what is taught and how it is taught gives the

evaluators insight into identifying weaknesses in instruction. For example, as Merrill

(2001) states that, "Learning is facilitated when the instruction demonstrates what is to be

learned rather than merely telling information about what is to be learned…" (p. 6).

Evaluation should focus on how instruction is designed to improve learning.

As another example of what evaluators should look for in their studies of

instruction, Merrill (2001) notes that student participation usually leads to better learning.

He states that; "Most of the current work in cognitive psychology has shown that students

learn better when engaged in solving problems" (p. 5). Reigeluth (1999) agrees; " If

someone wants to learn a skill, then demonstrations of the skill, generalities (or

explanations) about how to do it, and practice doing it, with feedback definitely make

learning easier and more successful" (p. 14).

Why conduct a metaevaluation? Although most evaluation studies do not include

formal metaevaluations, professional evaluators are encouraged to conduct their own

internal metaevaluations while planning and conducting their own primary evaluations

and to invite others to metaevaluate their final evaluation results to enhance interpretation

of their results. As Finn, Stevens, Stufflebeam, & Walberg (1997) note about a

metaevaluation they conducted of the Integrated Learning Systems Project evaluation in

New York Public Schools, "The meta-evaluation was designed and conducted and is

now reported to help these audiences assess the evaluation report and draw warranted

conclusions about the merit and worth of the learning systems project" (p. 159).

Metaevaluation, well done, can help stakeholders understand and trust primary evaluations. A team of representatives from organizations associated with many disciplines has collaborated for more than 25 years to create standards for judging evaluations. They have submitted their work to practicing evaluators and continue to obtain feedback for improving the standards in annual meetings. The results of their work are summarized in Program Evaluation Standards (Sanders, 1994) and are used in this project to guide the metaevaluation presented here because the intermediate course evaluation that was metaevaluated is considered a program evaluation. To summarize specific indicators regarding how well the standards are being met in a particular evaluation, Stufflebeam (2005) developed a checklist which also guided this metaevaluation.

<center>Methodology for the Metaevaluation</center>

As suggested in the literature review, the data collection methods for the metaevaluation focused on discovering how well the intermediate course evaluation met the thirty program evaluation standards and the particular list of indicators proposed by Stufflebeam (2005).

Data were collected through observation, review of reports, and interviews. The evaluators' procedures were observed and recorded in field notes identifying what they did to conduct the primary evaluation. Reports and evaluators' records were reviewed to clarify how they summarized the evaluation activities and findings. Conversations and informal interviews were conducted to understand the stakeholders' expectations and how well the evaluation process met their needs.

The metaevaluation was conducted throughout the time that the intermediate course evaluation was being conducted in an informal way. A more formal procedure was implemented after the final reports were completed. This procedure involved reviewing each of the thirty standards and identifying evidence from the metaevaluation data collection process to judge the strengths and weaknesses of the intermediate course evaluation in light of those standards.

In the following section, the primary evaluation of the intermediate course is presented to set a context for the metaevaluation which will follow in the subsequent and final section of this manuscript.

<div align="center">The Intermediate Course Evaluation</div>

In this section, to provide a context for the metaevaluation, the evaluand or object of evaluation (the intermediate course) for the intermediate course evaluation is identified. In addition, important background information, who the stakeholders were, their criteria for judging the evaluand, the questions answered, data collection and analysis procedures, reporting strategies, results, recommendations, schedule, and budget are summarized.

Evaluand

The evaluand was an intermediate second language course. The course description states that the course is an intermediate advanced grammar, reading, and culture class for native English speakers who want to develop their knowledge in this language. It introduces students with some experience in the language to culture, literature, and grammatical concerns through the reading of short stories, novels, and drama in the target language. This course is required for students who will take upper-

division courses for a minor or major in the target language and is required by the department to qualify students with non-university experience with the language to take a computer-based challenge exam that tests basic grammar knowledge. Depending on how well they perform on the challenge exam, students may obtain up to 16 credits for lower division courses.

Background

The course is offered at a western university that offers 40 to 45 second language courses regularly. The target course had not been evaluated for about 12 years. Therefore this evaluation and a subsequent plan to redesign the evaluand were established with the aspiration of retaining more students who would continue taking upper-division courses after taking the intermediate course. It was anticipated that this evaluation, followed by course redesign, would encourage more students to minor or major in the target language and better prepare them to compete for jobs.

The evaluation took place during the fall to prepare for course redesign, scheduled to take place during the subsequent winter. The instructors were to be trained at the end of summer and the new second language course was to be implemented in the Department of Languages the following fall. All this was part of an ongoing effort to ensure that this university would develop excellence in languages.

Stakeholders

To better understand the purpose and goal of the evaluation, stakeholders and their views were identified to include the humanities college, the targeted second language department, the faculty director assigned to coordinate all instruction associated with the second language course, instructors (four full-time professors and six native

speaking graduate student instructors), and students. These stakeholders were consulted throughout the evaluation process to understand the evaluand from each of their individual and collective perspectives, to clarify their views on what they felt the evaluand should be doing, how well it was serving them, and how it might change to better serve the students while meeting department and college expectations.

The faculty director was assigned to coordinate all instruction associated with the Second Language Course. In his first semester as director, he identified the need to evaluate the curriculum so that it could be improved through thorough redesign and materials development. He petitioned the assistance of external evaluators for this project and identified graduate students who had taken the necessary classes to qualify as evaluators and who had extensive experience in the targeted second language course.

There were 198 students enrolled in eight sections of the intermediate course which is the course being evaluated, and 60 students in subsequent upper-division courses who have completed the course being evaluated. Random samples from these groups of students were drawn for participation in the focus groups. Results from these samples were extrapolated to future students.

Criteria

The criteria for judging this evaluand were derived from interviews with stakeholders about their questions, concerns, and values associated with the evaluand. Two criteria were identified as essential for evaluating the intermediate course.

1. Students should be prepared to succeed in one or more advanced courses.

2. Students should successfully prepare themselves for the challenging exam.

Questions Regarding How Well the Evaluand Meets the Criteria

Based on the stakeholders' definition of the evaluand as the intermediate second language course and their criteria for judging it, the principal questions stakeholders wanted answered through this evaluation were:

1. How well does the course prepare students to take advanced courses in the target language?

2. How well does the course prepare students to take the challenge exam?

Data Collection Methods

Two types of data collection activities were used to document stakeholders' perceptions of the worth of the intermediate course curriculum in terms of discrepancies between program performance and their expectations for it: focus groups and interviews.

Focus groups. Focus groups with students from each of the eight sections of the intermediate course met to discuss their perceptions of the curriculum. Evaluators used the focus groups to follow up on the themes raised to explore issues, concerns, suggestions, and recommendations of the participants regarding the course. These discussions gave the students the opportunity to provide more details, examples, and stories to illustrate their answers to the questions regarding how well the evaluand met the criteria.

The design called for creation of five focus groups of eight students each through random selection of five students from each of the eight sections of the intermediate course. An additional three students were randomly selected from the remaining students in each section to be alternates in case any of the first five selected students could not meet at times scheduled for the focus groups. The selected students were given a memo

(See Appendix A for the Randomly Selected Focus Group Memo) inviting and encouraging them to sign up. By mixing the students from different sections together students in each focus group listened to and responded to the issues brought up by students from other sections.

The day before each focus group an email was sent out to remind students of the upcoming focus group they had signed up for and phone calls were made to encourage them to come participate (See Appendix B for the Focus Group Reminder Email). Pizza and drinks were provided to motivate students to participate. A total of 32 students (out of 40 invited), or 16% of the enrolled students in the intermediate course attended the five focus groups.

At the beginning of each focus group session, students were given a consent form to read and sign (See Appendix C). After that, everyone in the focus group sat on comfortable chairs or couches around a table in a quiet room to discuss a series of prepared questions (See Appendix D). These questions were derived from the principle questions regarding how well the evaluand met the criteria. They were designed to help students' share their perceptions of what the course should do and how well it met the criteria.

The focus groups lasted about 1 hour and 15 minutes on average. The students said they would not mind staying a little longer so they could completely express their ideas. After completing the focus groups, pizza and drinks were served, and the students continued to talk about the course and shared additional and vital information that was included in the data analysis.

Interviews. Interviews were conducted with the six native language-speaking graduate student instructors and the four full-time professors in the department. The interviews were conducted to obtain perceptions of those who teach the course and subsequent courses regarding the academic worth of the intermediate course. Specifically, these interviews were intended to elicit from graduate students and professors their different but complementary perspectives on the evaluand compared to how they felt it should perform.

Although two different sets of questions were developed for these interviews, one for graduate instructors (See Appendix E for the Interview Questions – Graduate Student Instructors) and the other for professors (See Appendix F for the Interview Questions – Professors), they both covered the same themes. The graduate student instructors were asked more questions than the professors because their questions dealt with the actual daily teaching of the intermediate course, which did not apply to the professors who teach subsequent classes to students who have already taken the intermediate class.

These 10 interviews, which were scheduled to last about 30 minutes each, averaged about an hour each, and were conducted in the offices of the graduate student instructors and the professors to provide a safe, familiar, and comfortable environment.

Data Analysis Procedures

Qualitative procedures were used to analyze the qualitative data from course focus groups and interviews. That is, the texts from these data collection sources were studied and analyzed in order to categorize the data into major themes associated with the evaluation questions. The details, descriptions, examples, and stories that resulted from

the qualitative data collection activities illustrate the stakeholders' perceptions of what the evaluand was and what they believed the evaluand should be.

Reporting Strategies

There were several reporting strategies for this evaluation. First, an interim report of ten pages was delivered to the director of the course to begin the curriculum design project. This interim document synthesized the results and general themes regarding stakeholders' perceptions of the course and their ideas regarding the curriculum design, coupled with the evaluators' recommendations. A later evaluation report was written and presented to the dean, the director, and faculty in the languages department, addressing their questions.

This separate final report is presented to the Department of Instructional Psychology and Technology to provide context for the metaevaluation which follows and will be used to satisfy requirements for the Masters degree.

Results of Intermediate Course Evaluation

Class covered too many topics. The students felt that this course attempted to cover too many topics. They expressed their displeasure in trying to study grammar, literature, and culture all in one class. As one student said; "I spent hours in subjects that were not on the exam; [the study of] literature was busy work, useless and a waste of time."

The students and director all described the course curriculum as a "hodge-podge" that may have emerged as such because the graduate student instructors were given a textbook a few days prior to the beginning of each semester and told to focus the course 1/3 on culture, 1/3 on literature, and 1/3 on grammar.

As a result students felt they learned very little about many things, all of which they quickly forgot because the course encouraged them to cram instead of learn. One student summarized his disappointment, "I did not gain any better understanding of the language." Faculty who teach upper level courses confirmed that the course has not adequately prepared all students equally and evenly for success in advanced courses.

Language proficiency unimproved. The most common theme expressed by the students regarding second language pedagogy was that they did not have enough opportunities to practice the language to be prepared for advanced classes. As one student said, "I don't feel prepared to take advanced classes." They expressed a strong desire to have more pair and group work doing collaborative and communicative language activities. Because of the lack of language practice and feedback in the class, many students felt that their language proficiency had not improved. Others felt that their language proficiency had actually declined from the level they had from previous immersion experiences with the languages, as a student explained; "I have forgotten much of what I learned."

Instructors did not use the syllabus. From the comments made by the students in the focus groups, it is apparent that none of the instructors used or followed the course syllabus, which was designed to help the students prepare for subsequent advanced courses. The majority of the students did not know what the upcoming assignments would be or when the examinations would be administered. One student expressed her concerns about it saying, "We need more focus on the course." Another said, "I did not know the objective of the course."

Instructors did not give feedback on students' performance. Additionally, the students generally complained about not getting feedback on their assignments. Worse than this, most of the instructors were perpetually behind in grading, as one student stated, "The teacher was behind on grading and was not worried about it." As a result, students did not know how well they were doing in the course or how they could improve in their preparations for subsequent courses.

TAs were not trained as teachers. Many of the students expressed concerns regarding their instructors' qualifications and preparation to teach. Among all six instructors, only one was well regarded by the students. In most cases, the students felt that without better preparation their instructor could not prepare them for advanced classes or to take the challenge exam. One student declared that, "The teacher confessed not knowing what was in the exam." Although the instructors are easily accessible and very personable, the students tended to complain about their ability to teach. Some blamed this on the instructors' personal preparation while others blamed lack of department support.

TAs did not prepare helpful lessons. A common theme expressed by the students regarding the instructional methodology was that many of the teachers had the students read aloud in turn from the textbook material that were assigned as homework the day before. This proved to be a very unproductive learning experience and suggested to the students that their instructors were not preparing classes sufficiently beforehand to inspire them to learn and prepare for subsequent courses. Another problem that the students voiced was that the principal instructional methodology was lecture. The students expressed boredom and frustration with this non-interactive instructional approach to

teaching. As one student noted, "It was too much reading from the book, it did not catch my attention."

Textbook was for native speakers. Almost all the students complained about the quality of the textbook. They reported that the textbook was written for native language-speaking students. As one student said, "The book was hard to understand." Another added, "The book assumes that the learner knows the historical and cultural facts behind the literature pieces and grammar examples." Another student claimed that; "Students were expected to try to explain the literature even though they did not understand it [grammatically or in terms of the literary quality]." Leaving so much up to the students led many of them to decide not to take upper level classes. As a student declared, "I don't want any other class on this language."

Also, the grammar explanations in the textbook were not very clear and the grammar examples were not very useful. A student reported; "The book was written for native speakers, it is too complex." The students also noted that there were not enough opportunities to practice grammar usage in the textbook, which focused more on historical facts and cultural issues. As a result, the students felt they were not prepared for the challenge exam.

Professors' views differed from others. Interestingly, three of the four professors were opposed to using this evaluation and subsequent curriculum revisions to recruit more students into upper-division courses or to declare a minor or major in the language. They were happy with the small class sizes in their upper-division grammar, literature, linguistics, and culture courses and did not want to add more sections of these classes with more students enrolled in each section.

Not enough grammar was taught. The students wanted to learn more grammar in preparation for the challenge exam than was given. One student summarized, "It would be more successful if the course taught us more grammar and writing and less culture." Another student said, "The course did not adequately prepare me to do well on that challenge exam." A third student in an advanced class who had the exam concluded, "Since the exam was only testing for grammar the [intermediate] class should teach grammar."

Course and exam relationship confused. Several students with extensive experience in the target language reported in interviews and focus groups that they were interested in taking the challenge exam to earn credits without having to take lower level classes. They were told by department secretaries that this was one important reason for taking the intermediate course. However, the official objectives of the course outlined in the course description did not list preparation for the challenge exam. Instead, the description alluded to this course being used as a portal course into upper-division courses associated with a minor or major. The discrepancy between the course description, the reasons department secretaries gave students regarding why they should take this course, and the students' expectations for the course was a major concern for these students.

No course focus on the challenge exam. Students were not advised regarding when the exam was going to be given or how to take the exam because the instructors did not have this information or any mandate to teach it. As one student from the advanced class declared, "This class failed to prepare me for the exam." And another agreed, "I felt

that signing up for this class was a waste because it did nothing to prepare me for the exam."

Recommendations from Intermediate Course Evaluation

The previous section outlined results from the focus groups and interviews. Based on these findings, the following recommendations are offered.

1. It appears that the intermediate language course is being offered as an appetizer to other upper-division grammar, literature, culture, and linguistics courses. Most students wanted this class to prepare them well for the challenge exam, while a few want to go on and take advanced classes. So, it would be profitable to the students if this course spent the first half of the semester reviewing and teaching grammar so students could be prepared to take and pass the challenge exam. That exam should be announced well in advance, simulated in class to help students become familiar with its procedures, then be offered at midterm. After the exam, the second half of the course could survey the topics covered in the other upper-division courses and highlight the benefits, including job opportunities, of majoring or minoring in the language. The survey should cover in depth a few well-selected topics rather than superficially introducing several miscellaneous ones.

2. There was a unanimous agreement among most of the students that TAs were not giving information about assignments in advance. Teacher could give students an advanced organizer, a preview of the material that they will read for homework through a schema building activity at the end of the class period. Then the students could go home and prepare that material for the next class. The next class

should be filled with practical application activities that help the learners develop a deeper understanding of the material. The most productive activities are ones that require interaction with others using the target material.

3. The textbook used in this language course does not seem appropriate for native English speakers at the intermediate level. A grammar packet could be prepared and required for the class. This packet might include a review of fundamental grammar as well as cover the grammar problems demonstrated by students who have some experience with the target language. This packet needs to have numerous examples and a lot of practice exercises. It might also include activities that promote interaction and collaboration between classmates.

4. Students complained about the lack of interaction, thus the second language pedagogy needs to be developed around the principle of interaction as explained previously by Merrill (2001). Therefore, the classroom could provide multiple opportunities in every class for each student to interact with each other. This is easily done by putting students in pairs or groups and giving them communicative tasks to complete that require the use of the target material. Another pair or group activity that could be used involves giving students a problem solving task that requires interaction using the target material.

5. Most students did not receive a syllabus for the class and were confused about expectations. A syllabus could be developed and presented to class early in the semester and could be used as a contract agreed upon by the instructor and the students to guide the course. It could list the responsibilities and expectations for the students in terms of well-defined assignments and assessments, as well as

clearly articulated evaluation criteria for judging student work. Additionally, it could serve as a table of contents and a calendar to help students and instructors prepare for course activities. A single course syllabus with clearly outlined objectives could be developed and uniformly implemented in all sections of the language course. Once this syllabus is presented to the various sections, it might be followed by both the instructors and the students.

6. Some students felt the instructors were not as prepared as they might have been. The instructors will fare much better if they are well prepared to teach the course. This could be improved in three ways. First, assignments to teach could be made well in advance so that instructors can prepare lesson plans and complementary materials. Second, the department could organize weekly instructor meetings during which instructors are taught effective teaching methodology. The weekly instructor meetings are also a good opportunity to share materials and resources. They also afford instructors opportunities to gain insights into how various topics can be more effectively taught. Third, instructors might be required to take a second language methodology course their first semester of teaching. However, it is recommended that such a course be completed before the first teaching assignment is made.

Schedule

Work on this evaluation study began when the director requested help in finding external evaluators. Two evaluators began to meet with the director the next week. The following weeks were filled with interviews with stakeholders, the focus groups, the teachers and instructors, through the first half of December. The last half of fall involved

data analysis. An interim report of preliminary findings and initial recommendations for the curriculum development stage of this project was due by end of fall; however, analysis, and reporting continued throughout the following year. The curriculum design project was to be done during winter based on the interim evaluation report submitted and presented at the end of fall. Table 1 outlines the planned compared to time spent for the major events of this evaluation. Differences between what was projected and what actually happened were due to the client deciding to change the project midway through implementation.

Table 1

Comparison of planned and time spent.

Task name	Start	Finish	planned	lasted
Conduct initial interviews with the client	11/5	11/14	5 days	8 days
Do literature review of second language acquisition curriculum evaluations	11/7	11/13	7 days	5 days
Write evaluation proposal and present it to the department of languages faculty.	11/11	11/13	3 days	3 days
Develop data collection instruments (focus group and interview questions).	11/14	11/25	8 days	8 days
Prepare for data collection (interviews and focus groups).	11/26	12/1	3 days	4 days
Conduct focus groups.	12/1	12/5	5 days	5 days
Conduct faculty and instructor interviews.	12/10	12/18	10 days	9 days
Do qualitative data analysis on focus groups and interviews.	12/12	12/26	11 days	11 days
Write interim evaluation report.	12/29	12/31	2 days	3 days
Write executive summary and final evaluation report.	1/2	1/23	13 days	16 days
Present findings to the Department of languages.	1/23	1/25	2 days	2 days

Budget

Table 2 reports the projected budget and the actual expenditures. In fall, the

evaluation team worked about 150 hours on data collection and data analysis. Writing

and presenting the evaluation reports required another 150 hours, totaling about 300 man

hours to complete the evaluation.

Table 2

Comparison of the projected budget with the actual expenditures.

Itemized Resource	Projected Budget	Actual Expenditures	Difference
Evaluation Staff	6,000	5,325	675
Consultants	220	0	220
Faculty	750	500	250
Secretaries	120	84	36
Printing & duplication	100	75	25
Data processing	75	75	0
Printed materials	25	10	15
Supplies and equipment	200	125	75
Subcontracts	25	25	0
Totals	7,515	6,219	1,296

Although this evaluation had a relatively large projected budget, most of these

costs were defrayed by taking advantage of available local resources that were already

paid for by the university. For example, the professors' and secretaries' time were paid by

their normal salaries. Additional costs were cut by using equipment, materials, and

supplies already available through the university. The consultants were not used because

there was too little time to involve them.

Summary of Intermediate Course Evaluation

As the director of the program declared his desire to find and solve course

deficiencies, he initiated an evaluation of the course curriculum, trying to understand the

program and its stakeholders and then improve the curriculum. He planned to implement

the results of this evaluation project as soon as possible, which accelerated the study activities. The analysis showed that the curriculum was in need of reform, or better it needed to be rebuilt. The evaluation focused on the extent to which students feel the intermediate course prepares them to take advanced courses in the target language and how well they feel the intermediate course prepares them to take the challenge exam.

The evaluation concluded that according to the students, the course is not achieving its objective of preparing them for subsequent courses. They feel the course is lacking many important elements to make it effective in helping them advance to the next level of courses. Students did not feel that it motivated them to consider going beyond this class to minor or major in the target language.

Likewise, the students felt that the intermediate course did not prepare them to take the challenge exam as well as they hoped it would. They didn't understand how to take the challenge exam nor did they feel prepared in grammar skills to do well on the exam.

The Metaevaluation

The evaluation of the intermediate language course described above was designed to meet the Program Evaluation Standards; therefore, this Masters project uses those standards as a basis for critiquing or metaevaluating that evaluation. The standards are divided into the following four categories: utility standards, feasibility standards, propriety standards, and accuracy standards. Within each category, the associated standards are described, along with related criteria, against which the performance of the evaluation team in completing the intermediate course evaluation is metaevaluated.

Utility Standards

"The utility standards are intended to ensure that an evaluation will serve the information needs of intended users" (Sanders, 1994, p. 23) through attention to seven standards as follows.

Stakeholder identification. "Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed" (Sanders, 1994, p. 23). According to Stufflebeam (2005), there are criteria for meeting this standard:

> (a) clearly identify the evaluation client; (b) engage leadership figures to identify other stakeholders; (c) consult stakeholders to identify their information needs; (d) ask stakeholders to identify other stakeholders; (e) arrange to involve stakeholders throughout the evaluation, consistent with the formal evaluation agreement; (f) keep the evaluation open to serve newly identified stakeholders. (p. 1)

The evaluation team identified the Director of the second language course as the client, as well as the language department professors, the instructors, and the students as the stakeholders. Through a sequence of meetings the criteria and the main questions to

be answered were identified along with other possible stakeholders interested in this program. The stakeholders were involved as much as possible, but some of them were not available as often as needed. As the evaluation evolved the evaluation team members were open to sharing it with new stakeholders, though none were identified.

Evaluator credibility. "The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance" (Sanders, 1994, p. 23), Stufflebeam (2005) notes the criteria for meeting this standard:

(a) engage competent evaluators; (b) engage evaluators whom the stakeholders trust; (c) engage evaluators who can address stakeholders' concerns; engage evaluators who are appropriately responsive to issues of gender, (d) socioeconomic status, race, and language and cultural differences; (e) help stakeholders understand and assess the evaluation plan and process; (f) attend appropriately to stakeholders' criticisms and suggestions. (p. 1)

Each of the evaluators on the team addressed and followed the evaluation standards and procedures as they interacted with the stakeholders throughout most of the project. However near the end of the year, one of the stakeholders lost trust in one of the evaluators, which led to the premature ending of the project. Nonetheless, the stakeholders' needs and concerns were still addressed throughout the evaluation. The evaluators considered all the differences which could exist among the respondents by not discriminating against anyone in any focus group or interview, and by accepting their views and opinions equally. The continuous update reports made during frequent meetings allowed the stakeholders to assess the information and processes used in the

project. Nothing was changed in how the results were obtained and interpreted based on the loss of trust mentioned earlier. All stakeholder criticisms and suggestions were accepted and used in the study.

Information scope and selection. "Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders" (Sanders, 1994, p. 23). Stufflebeam (2005) identified the related criteria for meeting this standard:

(a) assign priority to the most important questions; (b) allow flexibility for adding questions during the evaluation; (c) obtain sufficient information to address the stakeholders' most important evaluation questions; (d) obtain sufficient information to assess the program's merit; (e) obtain sufficient information to assess the program's worth; (f) and allocate the evaluation effort in accordance with the priorities assigned to the needed information. (p. 1)

During each evaluation team meeting, considerations were made about what questions to address in the evaluation. Besides several questions identified at the beginning, as the project continued other questions were considered. In fact, the two main questions addressed in the evaluation report presented above were added through this process. The results of the interviews and focus groups provided sufficient data to address the stakeholders' concerns and generate useful recommendations.

Although some of the professors felt the program did not need the evaluation, they released information needed for the team to conduct the project. Several additional questions were asked by the evaluation in addition to those addressed in the report. The two main questions were selected because there was sufficient data available to the

metaevaluator to use in analyzing the results associated with those questions. Other

reports may be generated by the participants when and if they make available the rest of

the data addressing additional questions.

    Values identification. "The perspectives, procedures, and rationale used to

interpret the findings should be carefully described, so that the bases for value judgments

are clear" (Sanders, 1994, p. 23). According to Stufflebeam (2005) there are criteria for

meeting this standard:

> (a) consider all relevant sources of values for interpreting evaluation findings,
>
> including societal needs, customer needs, pertinent laws, institutional mission,
>
> and program goals; (b) determine the appropriate party(s) to make the valuational
>
> interpretations; (c) provide a clear, defensible basis for value judgments;
>
> (d) distinguish appropriately among dimensions, weights, and cut scores on the
>
> involved values; (e) take into account the stakeholders' values; (f) as appropriate,
>
> present alternative interpretations based on conflicting but credible value bases.
>
> (p. 1)

    As mentioned earlier, other values were at work among the stakeholders in

addition to those used to define the two criteria used in the report (that the intermediate

course should prepare students for upper level classes and to pass the challenge exam).

This was because the evaluation team identified all relevant sources of values; but in

determining the appropriate parties to make valuational interpretations and stakeholders'

values to include, the team members were limited in the data results they could use. The

client, who was a key stakeholder, chose to withhold some of the data that would have

been used to address a wider set of values, based on his loss of trust in one of the

evaluation team members. The two main questions were based on the stakeholders'

values but additional values could have been included if the trust levels had been

maintained.

Report clarity. "Evaluation reports should clearly describe the program being

evaluated, including its context, and the purposes, procedures, and findings of the

evaluation, so that essential information is provided and easily understood" (Sanders,

1994, p. 24). Stufflebeam (2005) identifies the criteria for meeting this standard:

(a) issue one or more reports as appropriate, such as an executive summary, main

report, technical report, and oral presentation; (b) as appropriate, address the

special needs of the audiences, such as persons with limited English proficiency;

(c) focus reports on contracted questions and convey the essential information in

each report; (d) write and/or present the findings simply and directly; (e) employ

effective media for informing the different audiences; (f) use examples to help

audiences relate the findings to practical situations. (p. 2)

Different reports were written for the director of the course, other members of the

department, and readers of this report. There were no special needs audiences. As

indicated earlier, the report presented earlier in this metaevaluation has not focused on all

the contracted questions (though the other two reports did) because of limited access to

the data by the metaevaluator. The results were presented clearly and participants'

statements were quoted in the report to exemplify each of the points. Besides the written

reports, no other media were judged necessary to appropriately inform the audiences.

Report timeliness and dissemination. "Significant interim findings and evaluation

reports should be disseminated to intended users, so that they can be used in a timely

fashion" (Sanders, 1994, p. 24). Stufflebeam (2005) identifies the criteria for meeting this standard:

(a) in cooperation with the client, make special efforts to identify, reach, and inform all intended users; (b) make timely interim reports to intended users; (c) have timely exchanges with the pertinent audiences, e.g., the program's policy board, the program's staff, and the program's customers; (d) deliver the final report when it is needed; (e) as appropriate, issue press releases to the public media; (f) if allowed by the evaluation contract and as appropriate, make findings publicly available via such media as the Internet. (p. 2)

The evaluators did well in terms of timeliness but not so well in terms of dissemination. Oral reports and an interim report were given on time to the client. But based on that report and other interactions, the schedule did not allow the evaluators to share the full report with anyone else. This prevented the evaluators from sharing reports and having timely exchanges with all pertinent audiences, including the public. This metaevaluation, which includes a version of the report as a basis for the critique, constitutes another attempt to share some of the results. However, because the client distributed the report himself and the evaluators were not involved, the results were not fully available to the metaevaluator.

Evaluation impact. "Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased" (Sanders, 1994, p. 24). Stufflebeam (2005) notes the criteria for meeting this standard:

(a) as appropriate and feasible, keep audiences informed throughout the

evaluation; (b) forecast and serve potential uses of findings; (c) provide interim

reports; (d) supplement written reports with ongoing oral communication; (e) to

the extent appropriate, conduct feedback sessions to go over and apply findings;

(f) make arrangements to provide follow-up assistance in interpreting and

applying the findings. (p. 2)

Although the evaluation team provided interim reports and supplemented the

written reports with oral reports, because they lost contact with the client and other

stakeholders before the project was complete, the impact was less than hoped. Feedback

sessions and follow-up assistance to keep all audiences informed throughout the study

became impossible when the client asked the evaluation team members to give him all

the results and to discontinue participation in the evaluation.

Feasibility Standards

"The feasibility standards are intended to ensure that an evaluation will be

realistic, prudent, diplomatic, and frugal" (Sanders, 1994, p. 63) through attendance to

three standards as follow.

Practical procedures. "The evaluation procedures should be practical, to keep

disruption to a minimum while needed information is obtained" (Sanders, 1994, p. 63).

Stufflebeam (2005) identifies the criteria for meeting this standard:

(a) minimize disruption and data burden; (b) appoint competent staff and train

them as needed; (c) choose procedures in light of known resource and staff

qualifications constraints; (d) make a realistic schedule; (e) as feasible and

appropriate, engage locals to help conduct the evaluation; (f) as appropriate, make

evaluation procedures a part of routine events. (p. 2)

Disruptions during data collection were minimal by selecting interview and focus

group sites that were quiet and in areas familiar to the respondents and in harmony with

their schedule. Although focus groups and interviews are not routine, they were

conducted using evaluation team members who had experience with the language and

issues involved. An analysis of the schedule indicated that the study was conducted as

planned and was realistic. Evaluation staff who were more politically sensitive would

have made the project even more successful by maintaining trust with the client.

Political viability. "The evaluation should be planned and conducted with

anticipation of the different positions of various interest groups, so that their cooperation

may be obtained, and so that possible attempts by any of these groups to curtail

evaluation operations or to bias or misapply the results can be averted or counteracted"

(Sanders, 1994, p. 63). Stufflebeam (2005) clarified the criteria for meeting this standard:

(a) anticipate different positions of different interest groups; (b) be vigilant and

appropriately counteractive concerning pressures and actions designed to impede

or destroy the evaluation; (c) foster cooperation; (d) report divergent views; (e) as

possible, make constructive use of diverse political forces to achieve the

evaluation's purposes; (f) terminate any corrupted evaluation. (p. 3)

Although the reduction in trust between one team member and the client led to

early termination of the project and reduced cooperation between those participants, the

evaluation was still sensitive to a variety of stakeholders and their values. This allowed

the team to address positions of students, faculty, and instructors and to use their views in judging the quality of the intermediate course.

Cost effectiveness. "The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified" (Sanders, 1994, p. 63). Stufflebeam (2005) identified the criteria for meeting this standard: "be efficient; make use of in-kind services; inform decisions; foster program improvement; provide accountability information; generate new insights" (p. 3).

The project was less expensive than anticipated through the use of in-kind services and materials and by using graduate students as evaluators. But the effectiveness of the evaluation was unclear due to conflicts between an evaluation team member and the client. Although new insights were generated, it is not clear from the report or from information gathered by the metaevaluator whether or not the results were used to revise the course or inform other decisions about the evaluand. Also, records accounting for how the evaluation funds were spent were not available to the metaevaluator for review.

Propriety Standards

"The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results" (Sanders, 1994, p. 81) through attendance to eight standards as follows.

Service orientation. "Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants" (Sanders, 1994, p. 81). Stufflebeam (2005) identified the criteria for meeting this standard:

(a) assess program outcomes against targeted and nontargeted customers'

assessed needs; (b) help assure that the full range of rightful program beneficiaries

are served; (c) promote excellent service; (d) identify program strengths to build

on; identify program weaknesses to correct; (e) expose persistently harmful

practices. (p. 3)

The evaluation focused on the two main criteria the stakeholders said they valued

in judging the intermediate course and all relevant stakeholders or beneficiaries were

included. The intent was to help the program provide better service to students and to

correct weaknesses. Although harmful practices were not identified, the evaluation was

open to finding these if participants had reported them. Although there were probably

strengths of the program as well as weaknesses, these were not identified clearly in the

report.

Formal agreements. "Obligations of the formal parties to an evaluation (what is

to be done, how, by whom, when) should be agreed to in writing, so that these parties are

obligated to adhere to all conditions of the agreement or formally to renegotiate it"

(Sanders, 1994, p. 81). Stufflebeam (2005) noted several criteria to include in writing for

meeting this standard: "evaluation purpose and questions; audiences; editing; release of

reports; evaluation procedures and schedule; evaluation resources" (p. 3).

Although the agreements were summarized in writing in an evaluation proposal,

they were not presented in enough detail to address all these standards and issues. Also,

the project began to be implemented before the proposal was finally completed; so it did

not serve adequately as a contractual agreement.

Rights of human subjects. "Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects" (Sanders, 1994, p. 81). Stufflebeam (2005) identified the criteria for meeting this standard: "follow due process and uphold civil rights; understand participants' values; respect diversity; follow protocol; honor confidentiality/ anonymity agreements; minimize harmful consequences of the evaluation" (p. 4).

Concern for human subjects was a high priority in this evaluation project beginning with the application for IRB approval for the proposed study. The protocols set forth in that application were followed and the evaluation team did all they could to maximize the security and comfort of the participants. They were promised anonymity and the report meets that requirement.

Human interactions. "Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed." (Sanders, 1994, p. 81). Stufflebeam (2005) identified the criteria for meeting this standard:

(a) consistently relate to all stakeholders in a professional manner; (b) honor participants' privacy rights; (c) honor time commitments; (d) be sensitive to participants' diversity of values and cultural differences; (e) be evenly respectful in addressing different stakeholders; (f) do not ignore or help cover up any participant's incompetence, unethical behavior, fraud, waste, or abuse. (p. 4)

As with the rights of human subjects standard, this standard and associated criteria were met by following the protocol agreed to by the IRB review. There is no evidence gathered by the metaevaluation to indicate concern with human interactions.

Complete and fair assessment. "The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed." (Sanders, 1994, p. 82). Stufflebeam (2005) noted the criteria for meeting this standard:

(a) assess and report the program's strengths and weaknesses; (b) report on intended and unintended outcomes; (c) as appropriate, show how the program's strengths could be used to overcome its weaknesses; (d) appropriately address criticisms of the draft report; (e) acknowledge the final report's limitations; (f) estimate and report the effects of the evaluation's limitations on the overall judgment of the program. (p. 4)

Although submission to this metaevaluation indicates the evaluation team's willingness to acknowledge criticisms of earlier drafts of the report and the final report's limitations, this was not done soon enough to estimate the impact of the limitations, to identify strengths as well as weaknesses and use them to overcome weaknesses, or to identify unintended outcomes.

Disclosure of findings. "The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results" (Sanders, 1994, p. 82). Stufflebeam (2005) clarified the criteria for meeting this standard:

(a) clearly define the right-to-know audiences; (b) report relevant points of view of both supporters and critics of the program; (c) report balanced, informed conclusions and recommendations; report all findings in writing, except where

circumstances clearly dictate otherwise; (d) in reporting, adhere strictly to a code

of directness, openness, and completeness; (e) assure the reports reach their

audiences. (p. 4)

Although the reports were balanced and yielded informed conclusions and

recommendations in writing and according to codes of directness and openness, the

evaluation team reported only to the director of the program and not to all the audiences

and stakeholders. Others who had the right to know were identified but were not

addressed because the reports were given only to the director.

Conflict of interest. "Conflict of interest should be dealt with openly and honestly,

so that it does not compromise the evaluation processes and results" (Sanders, 1994, p.

82). Stufflebeam (2005) noted the criteria for meeting this standard:

(a) identify potential conflicts of interest early in the evaluation; (b) as appropriate

and feasible, engage multiple evaluators; (c) maintain evaluation records for

independent review; (d) if feasible, contract with the funding authority rather than

the funded program; (e) if feasible, have the lead internal evaluator report directly

to the chief executive officer; (f) engage uniquely qualified persons to participate

in the evaluation, even if they have a potential conflict of interest; but take steps

to counteract the conflict. (p. 4)

As designed, the evaluation used two external evaluators and one internal

evaluator to balance out potential conflicts of interest and engage people with different

points of view in conducting the study. Likewise, records were maintained; however,

only one person, the director of the program, ultimately had access to the majority of the

records. Potential conflicts of interest exist because the director was an internal evaluator

and not the chief executive officer or the funding authority.

Fiscal responsibility. "The evaluator's allocation and expenditure of resources

should reflect sound accountability procedures and otherwise be prudent and ethically

responsible, so that expenditures are accounted for and appropriate." (Sanders, 1994, p.

82). Stufflebeam (2005) clarified the criteria for meeting this standard:

> (a) specify and budget for expense items in advance; (b) keep the budget
>
> sufficiently flexible to permit appropriate reallocations to strengthen the
>
> evaluation; (c) maintain accurate records of sources of funding and expenditure
>
> and resulting evaluation services and products; (d) maintain adequate personnel
>
> records concerning job allocations and time spent on the evaluation project; (e) be
>
> frugal in expending evaluation resources; (f) as appropriate, include an
>
> expenditure summary as part of the public evaluation report. (p. 5)

A budget was planned at the beginning of the project with all relevant expenses

pertaining to the evaluation purposes delineated in it. The budget was followed closely

and no extra expenses were made. In fact, the team performed well using less than what

was expected. Expense receipts were given to the director to justify the costs but the

evaluation team did not keep records to include in the metaevaluation. Nevertheless, a

summary of what was spent is estimated in the evaluation report presented above.

Accuracy Standards

"The accuracy standards are intended to ensure that an evaluation will reveal and

convey technically adequate information about the features that determine worth or merit

of the program being evaluated" (Sanders, 1994, p. 125) through attendance to the 12 standards that follow.

Program documentation. "The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified" (Sanders, 1994, p. 125). Stufflebeam (2005) identified the criteria for meeting the standard above:

(a) collect descriptions of the intended program from various written sources and from the client and other key stakeholders; (b) maintain records from various sources of how the program operated; (c) analyze discrepancies between the various descriptions of how the program was intended to function; (d) analyze discrepancies between how the program was intended to operate and how it actually operated; (e) record the extent to which the program's goals changed over time; (f) produce a technical report that documents the program's operations and results. (p. 5)

Initial information regarding the intended course was provided by the director and later compared to other people's descriptions of how it was put into practice. Discrepancies were noted and analyzed in the reports. Since the evaluation was of short duration, changes in course goals were not noted.

Context analysis. "The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified" (Sanders, 1994, p. 125). Stufflebeam (2005) described the criteria for meeting this standard:

(a) describe the context's technical, social, political, organizational, and economic features; (b) maintain a log of unusual circumstances; (c) report those contextual influences that appeared to significantly influence the program and that might be

of interest to potential adopters; (d) estimate the effects of context on program outcomes; (e) identify and describe any critical competitors to this program that functioned at the same time and in the program's environment; (f) describe how people in the program's general area perceived the program's existence, importance, and quality. (p. 5)

Although the report described how people in the program's general area perceived the program's existence, importance, and quality, most contextual details were omitted to preserve the identity of the participants. This omission makes it difficult for others to judge the applicability of the results to their situations but it seemed essential for ethical reasons.

Described purposes and procedures. "The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed" (Sanders, 1994, p. 125). Stufflebeam (2005) identified the criteria for meeting the standard above:

(a) monitor and describe how the evaluation's purposes stay the same or change over time; (b) as appropriate, update evaluation procedures to accommodate changes in the evaluation's purposes; (c) record the actual evaluation procedures, as implemented; (d) when interpreting findings, take into account the extent to which the intended procedures were effectively executed; (e) describe the evaluation's purposes and procedures in the summary and full-length evaluation reports; (f) as feasible, engage independent evaluators to monitor and evaluate the evaluation's purposes and procedures. (p. 6)

All the criteria included in this standard were met but it should be noted that the

purposes changed and so only a subset of the methods used were included in the report

presented above. Part way through the study the focus shifted to an internal evaluation.

The metaevaluator did not have access to those internal evaluation products. The

evaluation report presented above and this metaevaluation are based only on data

available to the metaevaluator.

Defensible information sources. "The sources of information used in a program

evaluation should be described in enough detail, so that the adequacy of the information

can be assessed" (Sanders, 1994, p. 125). Stufflebeam (2005) described the criteria for

meeting this standard:

> (a) once validated, use pertinent, previously collected information; (b) as
>
> appropriate, employ a variety of data collection sources and methods;
>
> (c) document and report information sources; (d) document, justify, and report the
>
> means used to obtain information from each source; (e) include data collection
>
> instruments in a technical appendix to the evaluation report; (f) document and
>
> report any biasing features in the obtained information. (p. 6)

A succinct description of a variety of data sources was made in the evaluation

report, including the collection conditions and data analysis procedures. Focus group and

interview protocols are presented in the appendices but it is not clear that they were

validated for those uses. No biasing features of the obtained information were identified

so they were not documented or reported.

Valid information. "The information gathering procedures should be chosen or

developed and then implemented so that they will assure that the interpretation arrived at

is valid for the intended use" (Sanders, 1994, p. 126). Stufflebeam (2005) listed the criteria for meeting the standard above:

> (a) focus the evaluation on key questions; (b) assess and report what type of information each employed procedure acquires; (c) document how information from each procedure was scored, analyzed, and interpreted; (d) report and justify inferences singly and in combination; (e) assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions; (f) establish meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures. (p. 6)

During initial evaluation team meetings team members discussed the kinds of questions and how to collect associated information appropriately. Each piece of information gathered was focused on particular questions and associated criteria, established through discussions with stakeholders. This process appears to have allowed the evaluation team to meet the validity criteria.

Reliable information. "The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use" (Sanders, 1994, p. 126). Stufflebeam (2005) described the criteria for meeting the standard above:

> (a) identify and justify the type(s) and extent of reliability claimed; (b) as feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses; (c) in reporting reliability of an instrument, assess and report the factors that influenced the reliability, including the

characteristics of the examinees, the data collection conditions, and the evaluator's

biases; (d) check and report the consistency of scoring, categorization, and

coding; (e) train and calibrate scorers and analysts to produce consistent results;

(f) pilot test new instruments in order to identify and control sources of error.

(p. 6)

During weekly meetings, the evaluation team collaborated to choose and develop

the focus group and interview protocols. These are procedures that are commonly used to

invite students and faculty to express their opinions and have shown their consistency or

reliability in many settings across many different groups meeting at different times. This

form of reliability is difficult to calibrate with a statistic since no quantitative scores were

involved.

Systematic information. "The information collected, processed, and reported in an

evaluation should be systematically reviewed and any errors found should be corrected"

(Sanders, 1994, p. 126). Stufflebeam (2005) listed the criteria for meeting the standard

above:

(a) establish protocols and mechanisms for quality control of the evaluation

information; (b) verify data entry; proofread and verify data tables generated from

computer output or other means; (c) systematize and control storage of the

evaluation information; (d) strictly control access to the evaluation information

according to established protocols; (e) have data providers verify the data they

submitted. (p. 6)

After collecting focus group and interview data, team members examined each

others' notes compared to the recordings made. Unfortunately, the team did not give all

the information providers the chance to review what was recorded from these sessions because they did not believe that would be feasible.

Analysis of quantitative information. "Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered" (Sanders, 1994, p. 126). Stufflebeam (2005) explained the criteria for this standard:

> (a) whenever possible, begin by conducting preliminary exploratory analyses to assure the data's correctness and to gain a greater understanding of the data; (b) report limitations of each analytic procedure, including failure to meet assumptions; (c) employ multiple analytic procedures to check on consistency and replicability of findings; (d) examine variability as well as central tendencies; (e) identify and examine outliers, and verify their correctness; identify and analyze statistical interactions. (p. 7)

Although quantitative data were collected for this study, none of it was available to the metaevaluator, so it was not considered in this metaevaluation.

Analysis of qualitative information. "Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered" (Sanders, 1994, p. 126). Stufflebeam (2005) described the criteria for meeting the standard above:

> (a) define the boundaries of information to be used; (b) derive a set of categories that is sufficient to document, illuminate, and respond to the evaluation questions; (c) classify the obtained information into the validated analysis categories;

(d) verify the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders; (e) derive conclusions and recommendations, and demonstrate their meaningfulness; (f) report limitations of the referenced information, analyses, and inferences. (p. 7)

As documented in the evaluation report above, the data collection and analyses focused exclusively on qualitative data and the criteria presented for this standard were met.

Justified conclusions. "The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them" (Sanders, 1994, p. 126). Stufflebeam (2005) explained the criteria for meeting the standard above:

(a) limit conclusions to the applicable time periods, contexts, purposes, questions, and activities; (b) report alternative plausible conclusions and explain why other rival conclusions were rejected; (c) cite the information that supports each conclusion; (d) identify and report the program's side effects; (e) warn against making common misinterpretations; (f) whenever feasible and appropriate, obtain and address the results of a prerelease review of the draft evaluation report. (p. 7)

Although the results reported appear to be justified by the data gathered and reported, other interpretations and misinterpretations are always possible and the report did not address these or mention any search for side effects or unexpected outcomes.

Impartial reporting. "Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings" (Sanders, 1994, p. 126). Stufflebeam (2005) described the criteria for meeting this standard:

(a) engage the client to determine steps to ensure fair, impartial reports; (b)

safeguard reports from deliberate or inadvertent distortions; (c) as appropriate and

feasible, report perspectives of all stakeholder groups and, especially, opposing

views on the meaning of the findings; (d) as appropriate and feasible, add a new,

impartial evaluator late in the evaluation to help offset any bias the original

evaluators may have developed due to their prior judgments and

recommendations; (e) describe steps taken to control bias; (f) participate in public

presentations of the findings to help guard against and correct distortions by other

interested parties. (p. 7)

Because the study focus shifted to an internal evaluation, much of the data and

several reports were unavailable to the metaevaluator to use in judging how well the

evaluation met this standard. It is possible that some inadvertent distortions could have

been made and additional steps to control bias and guard against distortions should have

been made.

Metaevaluation. "The evaluation itself should be formatively and summatively

evaluated against these and other pertinent standards, so that its conduct is appropriately

guided and, on completion, stakeholders can closely examine its strengths and

weaknesses" (Sanders, 1994, p. 126). Stufflebeam (2005) listed the criteria for meeting

this final standard:

(a) budget appropriately and sufficiently for conducting an internal

metaevaluation and, as feasible, an external metaevaluation; (b) designate or

define the standards the evaluators used to guide and assess their evaluation;

(c) record the full range of information needed to judge the evaluation against the employed standards; (d) as feasible and appropriate, contract for an independent metaevaluation; (e) evaluate all important aspects of the evaluation, including the instrumentation, data collection, data handling, coding, analysis, synthesis, and reporting; (f) obtain and report both formative and summative metaevaluations to the right-to-know audiences. (p. 8)

Although the original evaluation did not include an independent or internal metaevaluation component to provide formative feedback while the evaluation was conducted, this summative metaevaluation provides useful information to the stakeholders. It is based on well known professional evaluation standards using all the information that was available to the metaevaluator.

Conclusions and Recommendations of Metaevaluation

Based on the judgments of the intermediate course evaluation against the standards summarized above, several changes could be made to make the evaluation more useful. First, the external and internal evaluation team members could have interacted more frequently and productively with each other and with all the stakeholders so their values and perspectives could be shared more fully in the study and so the study would be more politically viable and cost effective.

Second, the study could have been improved by using the proposal to create a formal agreement regarding how changes would be made if purposes shifted, noting the need to search for unintended outcomes, providing means for emphasizing strengths as well as weaknesses and how to use the strengths to overcome weaknesses, addressing

potential conflicts of interest better, and insuring procedures for collecting accurate information and presenting details in the reports associated with the evaluation.

Third, if a formative metaevaluation had been conducted while the evaluation was in process, formative feedback would have guided the evaluation team's efforts to improve their performance while there was still time to make adjustments in the many ways noted throughout this metaevaluation.

Finally, lessons learned through this summative metaevaluation about how the evaluation succeeded and mistakes that were made could help readers who conduct and consume evaluations in the future.

References

Brown, J. D. (1989). Language program evaluation: A synthesis of existing possibilities. In R. K. Johnson (Ed.), *The Second Language Curriculum* (pp. 222-241). Cambridge: Cambridge University Press.

Eisner, E. W. (1981). Using Professional Judgment. In R. S. Brandt (Eds.) Applied Strategies for Curriculum Evaluation (pp. 42-47). Alexandria, VA.

Lynch, B. K. (1996). Language Program Evaluation: Theory and Practice. New York, NY: Cambridge University Press.

Lynch, B. K. (2003). Language Assessment and Programme Evaluation. George Square Edinburgh: Edinburgh University Press.

Merrill, M.D. (2001). First Principle of Instruction. Submitted for Publication to educational Research & Development, Utah State University.

Reber, T. (2001). *Effective Teaching Behaviors and Attitudes as Perceived by Foreign Language Teachers*. Doctoral dissertation. University of Arizona, Tucson.

Reber-Bell, T. (in press). Behaviors and Attitudes of Effective Foreign Language Teachers: Results of a Questionnaire Study. *Foreign Language Annals*.

Reigeluth, C.M. (1999). Instructional-Design Theories and Models Vol. II. A new Paradigm of Instructional Theory. Indiana University: Lawrence Erlbaum Associated Publishers.

Rodgers, T.S. (1989). Syllabus design, curriculum development and policy determination. In R. K. Johnson (Ed.), *The Second Language Curriculum* (pp. 24-34). Cambridge: Cambridge University Press.

Sanders, J. R., chair of the Joint Committee on Standards for Educational Evaluation

    (1994). *The Program Evaluation Standards: How to Assess Evaluations of*

    Educational Programs (2nd ed.). Thousand Oaks, CA: Sage Publications.

    Retrieved October 19,2005, from

    http://www.wmich.edu/evalctr/jc/

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Thousand Oaks, CA: Sage

    Publications.

Shinkfield, A. J. & Stufflebeam, D. (1995). *Teacher Evaluation: Guide to Effective*

    Practice. Boston, MA: Kluwer Academic Publishers.

Stufflebeam, D. L. (2005). *Program Evaluations Metaevaluation Checklist (Based on the*

    Program Evaluation Standards). In The Evaluation Center. Retrieved October 20,

    2005, from: http://www.wmich.edu/evalctr/checklists/program_metaeval.htm

Worthen, B. R., Sanders, J. R., & Fitzpatrick, F. L. (1997). *Program Evaluation:*

    Alternative Approaches and Practical Guidelines (2nd ed.). New York, NY:

    Longman Publishers USA.

Appendix A

Randomly Selected Focus Group Memo

Dear: Time:

                                              Place:

You have been randomly chosen to participate in a focus group that the Languages department is conducting. Because we want to offer the best language instruction possible, we are taking a careful look at the intermediate language class this semester. We would like to request your help.

We realize that this is a busy time of the year. However, because you have been selected through a random process, it is very important that, if at all possible, you be a participant.

We have scheduled meetings on different days and at different times with the hope that one of them will work for you.

The focus group will last one hour. As a small incentive, we will have pizza and drinks for you when you come.

Appendix B

Intermediate Language Course Focus Group Reminder Email (sample)

We appreciate the time you took to complete the Intermediate language class in-class survey and questionnaire this week. As mentioned in class, you were randomly selected to participate in a focus group to discuss in greater depth your evaluation of the course curriculum. Because you were selected through a random process, it is very important that you attend the focus group meeting. Your comments, coupled with those of your peers, will enable the Department of Languages to provide the best language instruction possible. The focus group that you signed up to attend is:

Date: Thursday, 4 December

Time: 11:00 a.m. - Noon

Place: 54041 HB

Please come 5 minutes early so that we can start on time. We will provide pizza and drinks as a small incentive and token of our appreciation of your willingness to participate in this one-hour focus group meeting at this busy time of the year.

The intermediate language Section of the Language Department

Appendix C

Consent for a Research Subject

Introduction

      This study is being conducted by the Languages Department in order to evaluate the adequacy and quality of this intermediate language course curriculum. Among the students who are currently taking this intermediate language class, you were randomly selected to participate.

Procedure

      You have already completed the university evaluation form and the written survey. During this focus group, you will be asked questions whose responses will help us further assess the quality of this intermediate language class. Questions will focus on, among others, your expectations, the value of the course to you, how the course prepared you for other classes, and how the course can be improved. Your participation will be audio taped so that the moderator can focus on the thread of the discussion rather than having to write down notes.

Risks

      There are no risks involved in your participation in this focus group. If you do not wish to answer a question, you may decline to do so. The moderator will be sensitive to your needs.

Benefits

      By participating in this focus group, you will be helping the department further refine the curriculum of this intermediate language class and, by extension, will be helping future students who will take the course.

Confidentiality

At no point will you be asked to identify yourself. All answers will remain confidential and will only be reported as group data with no identifying information. All data, including tapes and transcriptions will be kept in a locked storage cabinet and only those directly involved with the research will have access to them. After the research is completed the tapes and the transcriptions will be destroyed.

Participation

Your participation in this focus group is voluntary. You have the right to withdraw or refuse to participate at any time without jeopardy to you at any level. Questions about the Research and of your Rights as a Participant

If you have any questions regarding this survey, please contact the director at 822-2176 or director@mwu.edu. If you do not feel comfortable asking the researcher, you may contact the IRB Chair, 822-5490, 5120B RB, chair@mwu.edu.

I have read, understood, and received a copy of the above consent and desire of my free will and volition to participate in this study.

Name (Print): _____

Date: _____

Signature: _____

Appendix D

Intermediate Language Course Focus Group Questions

1. Why did you take the intermediate language class?

2. What expectations did you have for the intermediate language class?

    2.1. Were they met?

        2.1.1. How?

        2.1.2. Why not?

3. What is the best or most helpful/useful part of the intermediate language class?

4. Why?

    4.1. The textbook

    4.2. The way the course was taught

    4.3. The instructor

    4.4. The way the course was organized

    4.5. Etc.

5. What is the worst or least helpful/useful part of the intermediate language class?

6. Why?

    6.1. The textbook

    6.2. The way the course was taught

    6.3. The instructor

    6.4. The way the course was organized

    6.5. Etc.

7. How well did the intermediate language class prepare you to:

    7.1. take more advanced upper-division classes in this language,

7.2. complete a minor in this language, and

7.3. complete a major in this language,

7.4. use this language in your professional careers?

8. How can the intermediate language class be improved to better prepare students

to:

8.1. take more advanced upper-division classes in this language,

8.2. complete a minor in this language, and

8.3. complete a major in this language,

8.4. use language in your professional careers?

9. Is the intermediate language class successful?

9.1. Why?

9.2. Why not?

Appendix E

Intermediate Language Course Interview Questions

Graduate Student Instructors

As you know, we have been hired to work with the director of this course to evaluate the intermediate language curriculum. We would like the information that we collect to be useful to people like yourself. At this stage, we're interested in learning about your perceptions of the intermediate language curriculum and what the evaluation can do for you. Could we begin by your sharing your current thoughts about the intermediate language curriculum and this evaluation?

1. What do you perceive as the *strengths* and *weaknesses* of students with previous experiences with this language who took the intermediate language class?

    1.1. What areas should the *instruction* and *practice* focus on to better serve their needs?

2. What is your *general perception* of the intermediate language curriculum?

    2.1. What do you think of it?

        2.1.1. What *do you like* about the intermediate language curriculum? Why?

        2.1.2. What *don't you like* about the intermediate language curriculum? Why?

3. What do you think the intermediate language curriculum is designed to do? That is, what do you *perceive* as the *purpose*, *goals*, *objectives*, or *guiding philosophy* of the Intermediate language curriculum?

    3.1. How well do you think the *instruction* and *activities* that you conduct in class lead to the achievement of these goals? Please explain your answer.

3.2. Do you think the *content* areas the Intermediate language curriculum

  addresses are important? Why or why not?

4. What do you think the Intermediate language curriculum *should be* designed to

  do? That is, what *should be* the *purpose*, *goals*, *objectives*, or *guiding philosophy*

  of the Intermediate language curriculum?

  4.1. What would you *keep the same*? Why?

  4.2. What would you *change*? Why?

  4.3. What *instructional activities* do you see as the *most critical* for achieving

    those goals?

5. Having taught Intermediate language, what are *your* expectations/objectives for

  it?

  5.1. Are they being met? How or how not?

6. How well does the Intermediate language course prepare students to:

    6.1.1. 1) do well on the challenging exam,

    6.1.2. 2) take more advanced upper-division classes in this language,

    6.1.3. 3) complete a minor/major in this language, and

    6.1.4. 4) use this language in their professional careers?

  6.2. How can it be *improved* to better prepare students to be successful in these

    four areas?

7. The Intermediate language course is currently divided into thirds—1/3 grammar,

  1/3 literature, and 1/3 culture. Is this the best organization for this course? Why or

  why not?

  7.1. If not, how would you recommend the course organization be *changed*?

8. What have you found to be the *most* helpful or useful part of the Intermediate language course for the returned missionaries? That is, what would you *keep* in the curriculum? Why?

   8.1. What have you found to be the *least* helpful or useful part of the Intermediate language course for the experienced students in this language?

      8.1.1. That is, what would your *change*? Why?

      8.1.2. What would you *remove*? Why?

9. Do you receive sufficient support from the Department of Languages, this language Section in particular, to be successful in teaching Intermediate language? Why or why not?

   9.1. What *additional support* would you recommend the department or section provide you?

10. Is the Intermediate language course successful? Why or why not?

10.1 If not, what changes would *you* implement to make Intermediate language more successful? Why?

Appendix F

Intermediate Language Course Interview Questions

Professors

As you know, we have been hired to work with the director of this course to evaluate the Intermediate language curriculum. We would like the information that we collect to be useful to people like yourself. At this stage, we're interested in learning about your perceptions of the Intermediate language curriculum and what the evaluation can do for you. Could we begin by your sharing your current thoughts about the Intermediate language curriculum and this evaluation?

1. How prepared are students to take your upper-division courses? What abilities are they *lacking* that they should develop in the Intermediate language course?

2. What is your *general perception* of the Intermediate language curriculum?

    2.1. What do you think of it?

        2.1.1. What *do you like* about the Intermediate language curriculum? Why?

        2.1.2. What *don't you like* about the Intermediate language curriculum? Why?

3. What do you think the Intermediate language curriculum is designed to do? That is, what do you *perceive* as the *purpose*, *goals*, *objectives*, or *guiding philosophy* of the Intermediate language curriculum?

    3.1. How well do you think the Intermediate language course is doing what it was designed to do?

    3.2. Do you agree with these purposes or philosophy? Please explain your answer.

4. What do you think the Intermediate language curriculum *should be* designed to do? That is, what *should be* the purpose, goals, objectives, or guiding philosophy of the Intermediate language curriculum?

   4.1. What would you *keep the same*? Why?

   4.2. What would you *change*? Why?

5. Experienced students in this language are required to take Intermediate language before taking your courses. What objectives *should* this class have in order to prepare students to perform well in your classes?

6. How well does the Intermediate language course prepare students to:

   6.1.1. 1) do well on the challenging exam,

   6.1.2. 2) take more advanced upper-division classes in this language,

   6.1.3. 3) complete a minor/major in this language, and

   6.1.4. 4) use this language in their professional careers?

   6.2. How can it be *improved* to better prepare students to be successful in these four areas?

7. The Intermediate language course is currently divided into thirds—1/3 grammar, 1/3 literature, and 1/3 culture. Is this the best organization for this course? Why or why not?

   7.1. If not, how would you recommend the course organization be *changed*?

8. Is the Intermediate language course successful? Why or why not?

   8.1. If not, what changes would *you* implement to make Intermediate language more successful? Why?