

A UTILIZATION FOCUSED APPROACH TO THE EVALUATION OF A WEB BASED HYBRID CONVERSATIONAL MANDARIN PROGRAM IN A NORTH AMERICAN UNIVERSITY

Yao Zhang Hill & Stephen Tschudi
University of Hawaii at Manoa , USA

Abstract

The web based beginning conversational Mandarin class was launched at the University of Hawaii at Manoa during the Fall 2005 semester . The courses require students daily asynchronous participation online with once a week face to face meetings with a tutor . The program provides both an opportunity and a challenge to show the effectiveness of technology mediated instruction . The evaluators , who were also the instructors in the program , undertook a utilization focused approach , advocated by Michael Quinn Patton (1997) , and conducted a formative program evaluation of this instructional innovation . The fundamental principle of utilization focused formative evaluation is to fully utilize the evaluation findings to understand and to improve the program by the primary intended users (PIUsers) of the evaluation . Following the proposal of Patton (1997) , the project underwent the following phases : (1) identifying the constituents and the PIUsers of the program based on its developmental stage ; (2) identifying primary intended uses (PIU) of the evaluation ; (3) identifying and prioritizing the evaluation foci to make the evaluation manageable in the semester time frame ; (4) stage one data collection using existing evaluative documents to answer the prioritized evaluation questions ; (5) taking improvement actions based on the preliminary evaluation findings ; (6) stage two data collection to answer the primary evaluation questions , as well as to evaluate the improvement actions ; (7) modifying improvement actions based on students feedback , instructors observations , and cost effectiveness analysis ; and (8) designing a new cycle of evaluation to investigate the new questions . Upon reflection , both instructors , being evaluators at the same time , strongly consider the utilization focused approach to be accessible , effective , and empowering .

Key words

utilization focused evaluation ; instructional innovation ; participatory evaluation

1 . Introduction

Recent years have seen great growth in the program evaluation profession in language education in the United States . In contrast with the earlier evaluation practices that focused almost exclusively on measurement of outcomes using experimental designs and quantitative studies (e.g. Chastain & Woerdehoff 1968 ; Keating 1963 ; Scherer & Wertheimer 1964 ; Smith 1970) , the recent shift in the evaluation profession has observed more focus on program process and sees evaluation as a valuable tool to inform program development (Dassier & Power 2001 ; Kiely & Rea-Dickins 2005 ; Norris 2004 ; Yang 2007) .

However , an essential issue has yet to be brought to the central light of the evaluation practice —

utility. Before the late 1970s, evaluation activities were often conducted as scientific research by scholars in academics. Striving for rigor in methodology often compromised the evaluation's usefulness. Stakeholders often found that the evaluation was either incomprehensible or irrelevant. Meanwhile, evaluation was usually conducted as an isolated activity apart from the program curriculum. There was often lack of communication and joint decision-making between the evaluation staff and the program staff during the evaluation process. Failure to use evaluation findings has become an increasingly prominent issue brought to the attention of the community of evaluators. This was reflected in the revised standards for evaluation by the Joint Committee in 1994. In the revised standards, utility was listed as the top standard, followed by feasibility, propriety, and accuracy (Joint Committee 1994).

2. Utilization focused evaluation

In the search for a theoretical framework best at providing a solution to this issue, I found Michael Quinn Patton's (Patton 1997) utilization focused evaluation (UFE) to be an excellent approach to follow. According to Patton, program evaluation is "the systematic collection of information about the activities, program effectiveness, and/or inform decisions about future programming" (Patton 1997). There are two central concepts to the UFE: primary intended users (PIUsers) and primary intended uses (PIU). All the UFE activities are directed by one fundamental principle: the evaluation should be done for and with PIUsers for achieving PIU.

Following the UFE's principle, the first step in an evaluation is to identify PIUsers. PIUsers are "a group or individuals who are in position to make decisions about the program, intend to use the evaluation findings to inform their future program actions, and are affected by the evaluation outcomes" (Watanabe, González-Lloret, Hill & Yang 2007). Usually, for a language program, PIUsers can be funders, administrators, teachers, students, and sometimes parents and media, depending on the degree of their influence to the program. After the identification of the PIUsers, negotiation of the purpose of the evaluation will take place. According to Patton (1997), an evaluation can serve three primary purposes: "rendering judgments, facilitating improvements, and/or generating knowledge". Among these three purposes, judgment-oriented evaluation is analogous to Scriven's (1967) summative evaluation, which is about judging the overall effectiveness of a program to make termination or continuation decisions. On the other hand, an improvement-oriented evaluation features what Scriven (1967) defined as formative evaluation, which focuses on the improvement and enhancement of the program. While both judgment-oriented and improvement-oriented evaluations expect actions taken upon evaluation findings, knowledge-oriented evaluations serve only the conceptual use of findings for the stakeholders to better understand the program, reduce uncertainty, gain "enlightenment", and generate theories (Patton 1997).

After identifying the purpose of the evaluation, evaluators will work with the PIUsers to decide on the PIU, generate evaluation questions, prioritize the questions, design evaluation methodology, collect data, and interpret the results.

Each step of the evaluation features active involvement of the participation of the PIUsers. This implies several unique characteristics of UFE. First, the evaluation is necessarily participatory, since the PIUsers and/or stakeholders need to participate in almost all major decision-making in the evaluation processes. The active engagement on the part of the PIUsers are likely to create the sense of ownership of the evaluation, and thereby to increase the chance for the evaluation findings to be utilized. The utilization of evaluation can bring about the sense of empowerment for all the participants in the evaluation, because they can see the actions taken upon their input. Needless to say, to engage stakeholders participation and to facilitate the decision-making process to reach consensus, evaluators need to be active in pushing the process forward, reactive to PIUsers input, and adaptive to the dynamic situational context. Evaluators should consider their roles as collaborator, consultant, and educator, rather than just as methodologist and data analyst. The democratic negotiation and decision-making will not only lead to clear and understandable outcome, but also educational and transformative experience to the PIUsers, who very likely will gradually evolve as evaluators themselves. All the processes and products of the evaluation, in the end, are meant to achieve one goal: maximally make the use of the evaluation findings.

Unfortunately, up to this date, there are only a couple of evaluations using a utilization focused approach in second language and foreign language education. Norris (2004) conducted a three-year evaluation of the validity of the placement assessment program at Georgetown University German Department using UFE. The evaluation process went through many cyclical processes engaging all the stakeholders in the department, from administrators, teachers, staff, to teaching assistants. The placement assessment went through multiple stages of improvement and contextual appropriate adjustments, each stage utilizing the evaluation findings from the previous one. The process of using the evaluation findings also helped to clarify the program's goals and objectives and the reexamination of the curriculum, which led to the later program-wide task-based curriculum innovation (Byrnes 2002; Byrnes & Sprang 2003). Another case exemplifying the application and influence of UFE is Yong (2007), who initiated a two-semester long formative evaluation of a new teacher training program in the English Language Institute (ELI) at the University of Hawaii at Manoa (UHM). Not only were the evaluation findings utilized to make improvements for the program, but the evaluation process also transformed the attitudes of the administrators towards program evaluation. They changed from being doubtful to active engagement to enthusiastically inviting more evaluations to be done for the other components in the ELI program.

Following the same UFE principles and processes, the project reported in this paper is a formative evaluation of a hybrid web-based Mandarin program in the U.S. In the following sections of the paper, we will first introduce our web-based Mandarin program. We will then describe the process during which we identified the PIUsers, PIU, evaluation purposes, evaluation questions, and prioritized evaluation foci. Since our evaluation took two stages, we will present the methodology and main findings for the first stage, followed by those of the second stage. In the end, we will share our reflections and insights gained through the UFE process.

3. Program context

The program in focus is the first year web-based beginning Conversational Mandarin program (CHN111 and CHN112). The program is part of the innovative web-based language instruction courseware development project BRIX. It has been supported by the National Language Resources Center (NFLRC) and the Language Learning Center (LLC) at the University of Hawaii at Manoa (UHM). Those two courses were requested by the East Asian Languages and Literature Department (EALL) of the College of Languages Linguistics Literature (LLL) under UHM to offer to UHM students from the Fall 2005 academic year.

The program features students' daily asynchronous online participation with once-a-week 50-minute face-to-face meetings with a tutor. Students are supposed to follow a daily schedule and work with course materials (e.g. textbooks, listening workbook, and course CDs), as well as online materials (e.g. listening exercises, quizzes). Speaking tasks are often assigned in the online forums, in which the instructor assigns tasks with sample recordings, and students are supposed to complete the tasks with their own recordings, using the built-in MP3 recorder in the courseware. The enrollment of the course is between six to twelve students. Despite the small enrollment, the course was designed to set an exemplary model as to how a beginning level conversational Mandarin course should be taught in an online format. The program also has the potential to be offered to different campuses and other universities.

The first author of this paper was the tutor when the courses were first offered, and the second author has been the course developer, administrator, as well as the instructor. This evaluation project was initiated from our concerns for the program. Meanwhile, all the important stakeholders ranging from the instructors, to the NFLRC and the College of LLL were interested in the success of these classes, as a demonstration of the success of technology-enhanced foreign language classes. Since the program was still under the initial implementation stages, an improvement-oriented formative evaluation was needed at the time of the investigation.

3.1 Primary Intended Users and Uses

The first author launched the project initially in Spring 2006 to fulfill a graduate course requirement. She was also the tutor in the program at the time. As the primary evaluator, she realized that the most important stakeholder in this program is the course developer, administrator, and instructor: Stephen

Tschudi, the second author. Due to his thorough engagement in the evaluation and utilization process, Tschudi essentially became an evaluator of the program himself.

The first step that we took in identifying the PIUsers was to identify all the important constituents of the program and to analyze the relationship between these constituents and the program. Figure 1 displays the result of our analysis.

Figure 1. The structure of the constituents of the beginning conversational Mandarin online class (CHN111 /112)

Students are in the center of this diagram because they are whom this program has been designed to serve. Teachers undoubtedly play a central role in the program as they can directly influence the function and direction of the program. Other stakeholders who have a vested interest in these classes are: the EALL Department, the NFLRC, the College of LLL. The success of this program may bring them more funding.

The layers of circles reflect the College's power structure hierarchy with two facilitating institutions, NFLRC and LLC. They have provided vital supports in terms of providing the instructor, the technician, the graduate assistants, and all the technology support. Every single constituent shown in this chart has direct or indirect power to decide the fate of CHN 111 and 112. They are, therefore, all important stakeholders of these classes as well.

However, as the program was still in its infant stage, there were a lot of unknowns. It would be most productive and efficient for the instructors ourselves to understand it better before communicating with the wider scope of stakeholders. Also because the instructors were most familiar with program and could most directly act upon the evaluation findings, we decided to limit the scope of our PIUsers to the teachers circle, with students as our main informants.

The primary intended uses of the evaluation were set to improve the courses and generate deeper understanding of how the program works.

3.2 Evaluation questions

To fulfill the evaluation purpose, there were many questions to be asked. However, to make the evaluation manageable and efficient, we prioritized our questions into three foci:

- (1) How satisfied are students with each component of the course?
- (2) What are classroom related and individual related motivation factors?
- (3) What are the students' needs?

All three of these questions may be subsumed under a single question: How well are the courses functioning?

We also considered other evaluation foci, such as the appropriateness of the course current goals and objectives, and the sufficiency of the assessment system, but we agreed that it was still too early at this developmental stage of the program to investigate these foci.

This paper reports the first two stages of the investigation into our evaluation foci. In the first stage, we examined the existing evaluative documents, i.e., course evaluation questionnaires and students interviews, from the Fall 2005 semester. We implemented ten improvement actions in response to the findings. In the second stage, we asked students to evaluate the course again and to provide feedback on the improvement actions taken throughout the Spring 2006 semester. Further actions were planned to respond to the findings thereafter. The following section of the paper will report the process for each stage in turn, with a detailed description of methods and main findings.

4. Stage one evaluation

4.1 Methodology and data collection

UFE should be rooted in the program context. This implies that the first step in the evaluation is to examine the existing documents or information that can help better understand the program. As instructors, we had the unique advantage of having the most close contact with the program. But full engagement with daily pedagogical activities may also obscure us from looking at the program in a bigger picture and with a more objective perspective. During the evaluation process, we needed to distance ourselves from the daily routine time to time to examine the program and our roles in the program from a broader and more objective perspective. To achieve this purpose, we examined two sources of information: the responses to the final course evaluation given to students enrolled in CHN111 in Fall 2005 semester and students interview data gathered for a motivation study during the same semester.

The first source of data was from the CHN 111 course final evaluation (see Appendix 1). The questionnaire used open ended question format to elicit students feedback on six areas: (1) the online instructor and the tutor, (2) the material, (3) online activities, (4) overall effectiveness, (5) overall strengths and weaknesses, and (6) suggestions for improvement. Six students answered the questionnaire. Two of them were undergraduate students, two were graduate students, and another two were working professionals associated with the UHM. Three of them were females and three were males. Students completed the questionnaire in the last day of face to face tutoring session in Fall 2005 semester on a PC computer and saved the filled evaluation with a non sense name to protect their identity.

The second source of information was obtained from the semi structured interview data from another research project of the first author, investigating the effect of students motivation and self regulation strategies on their online learning achievement (Hill 2006)

Three out of six students of CHN111 voluntarily participated in the interview. One of them was an undergraduate student, one was a graduate student, and another one was a working professional. One of them was male and the other two female. All of the students interviews were audio recorded and transcribed. Students opinions on each element of the course —instructors, material, tutoring session, online activities, and suggestions for improvement have been extracted for the purpose of this evaluation project.

4.2 Primary findings of stage one evaluation

Table 1 categorized students responses on the course evaluation questionnaires into comments on either strengths or weaknesses of the program on each of the course components: teachers, material, online activities, tutoring content and overall evaluation of the course. The frequency of a particular comment was presented in the parenthesis in the end, if the occurrence is more than one. Table 1 tells us that students opinions toward the instructors were overwhelmingly positive. They also held positive attitude toward the course overall. However, the opinions toward materials and online components were more mixed. Students seemed to be happy with the textbook and the accompanying CD and the design and the content of the online components. Their complains seemed to concentrate on the difficulty of the Listening Workbook, the lack of oral interaction opportunities, and the technical difficulties with the online components.

Table 1 . Summary of Students Responses on the Fall 2005 Final Course Evaluation

Aspect	Strength	Weakness
Teachers	Effective and helpful (6)	
Material	Textbook and CD (3)	— Listening Workbook (4) — Textbook lack of interactiveness
Online	— Interactive — Easy to use — Good design — Exercise useful	— Bug — Slow downloading — Easy to fall behind — Pretty easy or too hard
Tutoring	Interactions	— Short and not enough
Content	Vocab learning	— Grammar patterns — Not enough practice for speaking (3)
Overall	— Flexible — Good structures — Pace — Convenience — Adaptive to individual needs	

Table 2 summarized three students interview responses on their opinions toward the course . Since the students who participated in the interview ,also filled the final course evaluation ,we can see that this summary overlap to a great extend with the course evaluation results . Some new information gained from the interview data include :(1) the need to practice longer speaking tasks ;(2) the need for the opportunity to review and repeat .

Table 2 . Summary of Students Opinions toward the Course Components from the Interview Data

Aspect	Strength	Weakness
Teachers	Tutor :good teaching style Online instructor : —good in giving feedbacks —experienced	
Material	Textbook : —build up vocabulary —easier and helps build up confidence Listening Lab Workbook :more interactive	Listening Lab Workbook —too hard (3) —no answer keys The textbook : —not very interactive —no answer keys — missing words in the gloss index
Online activities	Listening exercise :very interactive (2) Quiz :good Feedback on students recordings useful (2)	Recording : —not interactive (2) —too short Course software : —bugs —slow downloading
Tutoring	Favorite part Good to go to class (3)	Too short and not enough interaction (3)
Overall	More freedom and flexibility with time Encouraging independence Structure to follow	No need to study everyday Not enough reviewing and repeating Not enough speaking No synchronous response

We consider the students input extremely valuable to examine the potential weaknesses in the program . Meanwhile , we also realized that not all of the negative feedback reflected the inherent weakness of this online course in particular . Some of the problems may exist in the Mandarin program in general . Take the complaint on the lack of interaction opportunities for example . A closer examination in the responses to the interview question asking students to compare this online course with the brick and - mortar classes that they took before revealed a different story . Two students mentioned that , in this online course , they actually learned how to talk to people in Mandarin . One student mentioned that she felt more comfortable talking in Mandarin . Take the comments on the difficulty of the Listening Workbook for another example . Both instructors felt that the materials and exercises in the listening book were excellent . The conversations sounded natural and authentic and there was a gradual increase in difficulty in the exercises . It is a nice preparation for students to deal with real life communication situations . The tutor , also the first author , discussed this problem with students in class , and realized that the problems mainly lay in the lack of listening strategies . Many of the exercises only require students to understand keywords , as in the real life situations , but students always thought that they had to be able to recognize each sound and each word , which might be a difficult task even for advanced learners . Some other problems reported reflect the inadaptability of students in the online learning environment (e. g. “ no need to study everyday ” , “ easy to fall behind ”) , which might be ameliorated with more instructor guidance .

Despite the nature of the problems that caused students dissatisfaction , we decided to act upon our preliminary evaluation findings that related to related students learning needs to reducing demotivating factors . To summarize , students seem to :

- (1) have a desire for more speaking opportunities
- (2) want supplementary materials to increase opportunity and convenience for self study
- (3) be interested in more meaning based and discourse level tasks , and
- (4) need guidance to use learning strategies and to be autonomous learners .

One note of caution when interpreting the data is that the interview was not conducted with the sole purpose of course evaluation , and the Fall 2005 semester final course evaluation was developed in a short period of time without being piloted . They are able to give a general picture of students opinions and satisfactions , but they are only indirect sources of data . The conclusions are only tentative . The reason for the primary intended users to decide on using this information is that we wanted to take immediate actions to improve the course .

4.3 Acting upon primary evaluation findings

We focused on our negative evaluation findings to act upon in the hope to achieve maximum improvement effect . We decided to seriously take into account the students recommendations on how to improve the course . The main data source used is from students interviews because all of the students input were clarified by the first author during the interview , leaving little ambiguity as to what students wanted . Table 3 is a summary of students recommendations .

Table 3 . Students Recommendations on How to Improve the Course CHN111 and CHN112

Aspect	Recommendations
Material	<ul style="list-style-type: none"> 1 . Put lab book exercise online 2 . Balance difficulty level between materials 3 . Like to have characters 4 . Attach answers for both books 5 . Recommend a dictionary to accompany the textbook
Online	<ul style="list-style-type: none"> 1 . Online exercise downloadable 2 . Integrate speaking with listening —respond after listening 3 . Realtime conference 4 . Schedule outside classroom conference 5 . More on writing in paragraph 6 . Practice structure besides vocabulary

Aspect	Recommendations
Tutoring	<ol style="list-style-type: none"> 1 . More days to meet 2 . Summarize common mistakes 3 . Review the learned content 4 . Call upon individual to speak 5 . Each session longer
Overall	<ol style="list-style-type: none"> 1 . Identify each students weakness and suggestions to improve 2 . More practice on holding a conversation

Not all students recommendations are logistically possible to implement . For example ,in the category of material improvement ,we could not upload the textbook online because that would violate copyright ,in addition ,students already have a CD with the electronic form of the textbook . We also could not introduce Chinese characters due to the conversational nature of the course and the course tradition .In the tutoring category ,we could not change the college determined official schedule in response to the request for more days to meet and having longer sessions each time . We explained to the students why some actions could not be taken in response to their recommendations and students responded with understanding .

All in all ,we took ten improvement actions based upon our findings ,and they are :

- (1) 15 -minute conversation practice — The tutor met with students outside of the class and chatted about the topics that were previously covered about personal facts (e. g. name ,address ,hobby ,nationality ,schedule ,appointment) .
- (2) 5 -minute chat in tutoring — The tutor arranged the first 5 minutes of the 50 -minute tutoring session for students to carry out free conversations .
- (3) Assign paragraph writing online — Both instructors tried to give students communicative tasks , which require discourse level rather than sentence level or word level production . The instructors also provided sample writings and recordings . Sample tasks could be to describe the apartment that you ,a rental agent ,want to rent ,including the description of the furniture . Another example is to ask students to describe an imaginary dinner ,including the information about with who ,what ordered and what the prices are .
- (4) Summarization of common mistakes — All students online written and speaking responses (in Pinyin ,the Romanization of the Chinese sound system) were analyzed with problematic sentences identified and categorized by the first author . The second author compiled a summary of the common students mistakes and linked each category to the structure exercises designed throughout the semester to help students with these problematic areas . The document served as a review and was uploaded to the course website by the end of the course .
- (5) Scanning and uploading the answer key for the Listening lab workbook .
- (6) Checking textbook exercise online — Textbook exercises were checked and correct answers were provided through online forum called “ Language Clinic ” .
- (7) Individual diagnostic feedback — The first author identified individual students strengths and weaknesses in the areas of pronunciation , paragraph writing and recording , overall performance ,and learning strategies by studying their written and oral productions . A sample feedback form is attached as Appendix 2 .
- (8) Recommendation of dictionaries in both online and in book forms — we recommended other learning resources to students ,especially dictionaries in both online and book forms .
- (9) Review of postings during tutoring — Students online postings and recordings were reviewed during the tutoring time to time to reinforce learning .
- (10) Scheduling of office hour for tutoring — Students were encouraged to make appointments with both the tutor and the instructor to further practice using the language .

In addition to these formal actions taken ,both instructors also guided students to use learning strategies to approach different components of the course ,like the listening material in the Listening Lab Workbook .

5 . Stage two evaluation

All the improvement actions were taken toward the end of the Spring 2006 semester . Therefore ,we did not expect dramatic change in students attitudes toward each component of the course . To gauge the impact of these efforts we distributed a course evaluation on the last face to face meeting in Spring 2006 , specifically ask students to provide feedback on those actions .In addition ,both instructors also conferred on the cost effectiveness of these activities based the information they acquired from the informal communication with students and from their observations . Adopting ,adapting ,and abandoning decisions were made upon these improvement efforts based on our analysis . We also drafted future action plans to conduct needs analyses .

5 .1 Methodology and data collection

Three sources of information were collected during this phase of the evaluation ,incorporating course evaluation questionnaires as well as instructors observations and communications with students .

The Spring 2006 semester final course evaluation was a questionnaire consisting of both Likert scale questions and open ended questions . The questionnaires covered four major areas :(1) evaluation of the course components ;(2) evaluation of the improvement efforts ;(3)students needs analysis on topics of interest ;and (4) students self evaluation . The questionnaire is attached in Appendix 3 . Six students completed the survey ,three undergraduates ,one graduate student and two professionals . Four of them are females and two males . However ,not all students answered all questions . The questionnaire was given to students on the last tutoring session in as a Word document for them to type in the responses on PC computers . Students were asked to save their completed evaluation using a non sense name so that their identity could be anonymous to the instructors .

In addition to this formal data collection method ,we also made use of observations of the students reaction to our actions and informal communications with the students during the tutoring sessions and during the outside class conversational practice sessions .

5 .2 Primary findings of the stage two evaluation

In this session ,we will first report the questionnaire results on the evaluation of the components of the course and their evaluation of our improvement actions . We will then briefly discuss the findings on students topic interests and students self evaluations .

Table 4 is a collapsed summary of students numerical rating on the elements of each component of the course . K stands for the number of items /elements evaluated under each component category . The maximum rating is 10 and the minimum is 1 .

Table 4 . Summary of Students Average Evaluation Rating on Each Component of the CHN112

	K	M	SD
Instructor	7	9 .93	.09
Tutor	7	9 .88	.16
Material	4	8 .33	.30
Online	10	8 .88	1 .01
Tutoring	4	8 .55	.53
Overall	7	8 .77	.66

Again ,both instructors received the highest and most homogeneous rating on their teaching ,judged by the high average ratings (9 .93 and 9 .88) and small standard deviations (.09 and .16) . Materials received a relatively low average rating (8 .33) ,again due to the feeling of difficulty of the listening workbook (e. g. “the last part of each section is very frustrating ”) ,although some students were happy with them ,as one comment says ,“I like it the way it is . Both workbooks are clear ,helpful and move at just the right pace ” . The areas need to improve can be encapsulated into three themes :more practice of natural conversation , material topic and difficulty inappropriateness ,and lack of adaptation to online

learning . These issues will be addressed shortly . The findings for the evaluation of the improvement efforts will be presented next .

Seven improvement actions were presented to the students and Table 5 summed up the numerical results . The maximum rating is 4 (very useful) and the minimum is 1 (completely useless) . N is the number of respondents . One of the six students did not see the common mistakes summary and the uploaded answer keys by the end of the semester , while all others who utilized these materials found them “ very useful ” . Paragraph writing practice was well received as well . Students said it helped to think in Chinese and get the sentence structure right . It also seemed to make longer conversation easier to produce . One student found it “ tedious ” . The difference in the responses may due to students individual differences , such as onset language levels and different learning styles . Our two conversation oriented activities : 15 -minute conversation practice and 5 -minute chat received overall positive comments . Four students thought the 15 -minute conversation practice were great and desired more frequent practices . Only one student thought it was stressful . Three comments said that the 5 -minute chat was fun and useful , but one student said that some students were often late for the class . The individual feedback received decent rating as well ($M = 3.4$) , but one student commented that it was “ good to have but difficult to implement needed improvements ” . No comments was given for doing the textbook structure exercise online and on average students found it useful ($M = 3.2$) .

Table 5 . Summary of Students Evaluative Ratings on the Improvement Actions

Improvement Actions	N	M	SD
15 -minute conversation practice	6	3.4	.55
Individual Feedback Sheet	6	3.4	.55
Common mistakes summary	5	4	0
Uploaded answer keys	5	4	0
Paragraph writing practice	6	3.8	.45
5 -minute chat every beginning of the class	6	3.5	.55
Structure exercise online	6	3.2	.84

Three students who took CHN111 in the Fall 2005 semester , found CHN112 more satisfying . One student who newly jointed CHN112 in Spring 2006 felt himself more satisfied comparing to the previous Mandarin classes that he took .

Overall students attitude toward our improvement effort was very positive and encouraging . We received comments like “ I think this is a terrific class . I love that both teachers are so open to creating the environment that works best for the students . It's really shown this semester in the improvements that have been made , and it really made a difference ” . Again , the central overall concern is the lack of more natural conversation practice . The decision -making based on these feedbacks will be presented in the later section .

Next , we are going to briefly point out the highlights in the results of students opinions on useful topics and their self evaluation . The topics covered in the textbook perceived most useful by the students are : shopping , directions , and restaurant . Some outside textbook useful topics that were also suitable for this level of students are : telephone vocabulary , holiday vocabulary , sports or movies , extended food related vocabulary , bargaining , traveling , and describing spaces . Textbook topics hobbies , home , and school were considered less useful , one student even commented them as “ im material ” . Students felt a strong need to practice survival Mandarin when traveling to China , as one student recommended : “ basic things I would need to get by in China , and practicing this a lot ! ”

In the students self evaluation , on average , they deemed statements like “ I am really interested in learning Chinese ” and “ I seriously want to learn Chinese well in this class ” very true of themselves (both $M = 9.80$ out of 10 , $SD = .44$) and “ I am good at finishing tasks on time in this class ” the least true of themselves ($M = 4.80$, $SD = 3.03$) . A contrast like this indicated that even for highly interested and serious students , they might still have some trouble being timely in completing the online tasks .

Connecting all the information gathered in the evaluation survey ,overall ,it seems that the teachers effort in course improvement was recognized by the students and the students attitude toward the course and the improvement actions were very positive .

5.3 Acting upon stage two evaluation findings

There are two types of actions that we engaged in response to the evaluation findings . First we made adoption ,adaption ,and abandonment decisions on our improvement effects . Second ,we planned future actions to respond to the dominant findings on our evaluation . These decisions were based both on the survey and on our observation and informal communication with students .

About improvement actions ,there are three actions we did not put on the survey but decided to either give up or keep based on our observation . For example ,we abandoned reviewing online postings in the tutoring session ,because it could take up to half of the tutoring session ,which should be preserved for providing oral interaction opportunities . Instead ,we tried to develop more online tasks to increase language practice opportunity . We also did not elicit response to the action setting up office hour for language tutoring because no one took advantage of this service . However ,we notified students that they should feel free to make appointments with either instructor anytime during the semester . We also decided to keep recommending dictionaries to students because we considered it facilitative in students autonomous learning and it is very easy to do .

Similarly we decided to make the follow items permanent part of the course :a summary of students common mistakes ,answer keys for the listening workbook and checking textbook exercises in online " Language Clinic " ,where linguistic forms are specifically dealt with . These are examples of adoption decisions .

The 5 -minute chat in tutoring was considered successful as well judged from the survey ,but late comers ,even just one or two ,were very disruptive to the activity from the tutors observation . Also because the whole tutoring was designed to maximize interaction opportunity ,we decided to leave the continuation decision to the emerging situation nexttime the course offered ,such as enrollment ,students comfort level and learning needs .

Another oral practice oriented activity — 15 -minute outside class conversational practice — was welcomed by the students . We originally decided to conduct it four times in a new semester . However ,due to the increase in the enrollment and the cut of the funding for the tutor ,the online instructor's pedagogical duty was too heavy to carry out this activity . Instead ,the instructor developed much more interactive online activities where students have to listen to and respond to the previous recordings in order to carry on with their production . This decision is representative of an adaption type decision .

Also adapted was the paragraph writing activity .It was considered very useful by students but from the observation of and chat with the students ,we knew that very often these tasks would take long time for students to complete ,having to look up the right pinyin spelling and the tone marks . We decided to sequence the task complexity more gradually and require less "transcription " of what students need to record .

As for the individual feedback ,although it was rated very useful by students ($M = 3.4$ and $SD = 0.55$) ,we were doubtful how much students utilize the information and took actions to deal with suggested improvement areas . We thought it would be more cost efficient to provide ongoing individual feedback on a daily basis in the form of feedback to their task performance .

The aforementioned decision making was about whether and how to continue to carry out our ten improvement actions based on their impact and cost effectiveness . In response to the evaluation information on the course ,as summarized in the evaluation findings section ,we realized that there were three main areas we need to focus our improvement efforts . The first ,and the most important one ,is to increase the practice of more natural conversation and cover topics that are considered useful to students . To respond to this problem ,we took several actions . For example ,we decided to adjust the pedagogical weight of different units covered in the textbook and spend more time in teaching what students considered to be more useful to them ,including shopping ,food ,and directions . We significantly

modified the units that students thought irrelevant to their needs. For instance, in the unit teaching common vocabulary about school objects, instead of teaching vocabulary like roller pen and paper waste basket, we decided to ask students to provide vocabulary that are more relevant to their daily life, such as cell phone, laptop, keys, wallet, etc. However, we think these actions are still insufficient. Our determination to teach students what they need to learn subsequently led to a more comprehensive examination of students' needs, which later set the curriculum innovation in motion toward a task-based curriculum development. These efforts were reported in Hill and Tschudi (2007).

The second issue is related to material: the appropriateness of the topics and the difficulty level. This is closely associated with the first issue just discussed. We thought creating our own materials might be necessary, incorporating authentic conversations to cover topics of interest. Efforts were taken in spring and summer 2007 to collect authentic speech samples on shopping, directions and transportation in both Hawaii and Mainland China. Some materials were already made upon the analysis of some speech samples. More materials are expected to be created based on further discourse analysis.

The third issue concerns students' (lack of) learning strategy in the online learning. While still searching for better solutions to enhance online learning, we tried to first raise the awareness in the students that they need to change their learning style and develop strategies to suit online learning. We did this by sending out an email to all new incoming students a letter composed of the suggestions from previous students based on their online learning experience. The instructor also reminded students repeatedly as how to use the online schedule in the hope to improve their time management.

To sum up, we took all the actions that were possible and appropriate to our situation context under the constraints of time and resources in response to major evaluation findings. Our efforts in general helped to improve the most needed areas for improvement and they were greeted with warm welcome from the students.

6. Conclusion and reflection

During the process of utilization oriented evaluation, we found this evaluation approach extremely effective and fruitful. It helped us to successfully achieve our evaluation purposes: to improve and to generate deeper understanding of the program. We found out that students held overall a very positive opinion toward the program, especially toward the instructors. We identified four main areas for improvement: more practice of natural conversation, increasing relevance, appropriateness of materials in both topics and difficulty level, and insufficient use of learning strategies for online learning to achieve autonomous learning.

However, considering that the program was only in the very beginning stage of development and all students were new to this format of instruction, the existence of these areas of concern was very understandable. We felt strongly our evaluation matches what Patton (1997) calls the inventory checking stage of program evaluation. Just as one will not evaluate the function of a new hotel before beds and chairs are put inside, it would also be unfair to summatively evaluate an emerging program without first making sure it is equipped with all the necessary resources for the program to function. In our case, the evaluation led us to request answer keys and transcripts for the listening exercise book, only to discover that the transcripts were not published yet.

Even though this evaluation project is small in scale, some lessons learned and insights gained could be applied to any evaluation practice. It is pleasantly surprising to both of the evaluators how transformative and empowering the evaluation can be. The second author was at first just the primary intended user of the evaluation, but gradually he evolved into an evaluator himself, actively participating in the planning and decision-making process. The effect of the improvement that we brought about to the program made us see the power that we have to change the program and for students to feel empowered as well, because all our actions were based on their feedback of the program.

Several factors made it possible for our evaluation to achieve such an effect. First, we focused on our local context and maximizing the existing sources of information, and our evaluation went through an evaluate-react-reevaluate-react cyclical process. The cycle gave momentum to the evaluation practice and

it motivated the primary intended users to make more improvement efforts since we could see the changes ,and we got positive feedback from students .

Second ,the primary evaluator was active ,reactive ,and adaptive in the process ,maximally engaging the PIUser . While ,on the other hand ,the PIuser ,the online instructor of CHN 111 and 112 ,is an active person ,enthusiastic teacher ,and productive researcher himself . He is very cooperative and eager to make improvements to the course . The good personal communication helped to speed up the whole process as well .

Third ,following the UFU s framework made it possible to achieve the highest level of utilization of the evaluation findings .

Overall ,the process demonstrated its appropriateness and success in the formative evaluation of the beginning level Mandarin conversational program .

References

- Byrnes ,H . 2002 . The role of task and task based assessment in a content oriented collegiate foreign language curriculum . *Language Testing* 19 /4 :419-437 .
- Byrnes ,H . &K . A . Sprang . 2003 . Fostering advanced L2 literacy :A genre based ,cognitive approach . In H . Maxim &H . Byrnes (eds .) . *Advanced Foreign Language Learning :A Challenge to College Programs* , pp . 47-85 . Boston :Thomson -Heinle .
- Chastain ,K . D . &F . J . Woerdelhoff . 1968 . A methodological study comparing the audio lingual habit theory and the cognitive code learning theory . *Modern Language Journal* 52 /5 :268-279 .
- Dassier ,J . P . &W . Power . 2001 . Formative foreign language program evaluation :Dare to find out how good you really are . *Dimensions* 2001 :The Odyssey Continues . Selected Proceedings of the 2001 Conference of the Southern Conference on Language Teaching ,Birmingham ,AL ,15-30 .
- Hill ,Y . Z . 2005 . The effect of motivation and self regulation on online language learning achievement :A preliminary research . Honolulu ,HI :Unpublished term paper at the University of Hawaii of Manoa , Department of Second Language Studies .
- Hill ,Y . Z . &S . Tschudi . 2007 ,September . Exploring task based curriculum development in a hybrid web-based conversational chinese program . Paper presented at the 2nd International Conference on Task-Based Language Teaching ,Honolulu ,HI .
- Keating ,R . F . 1963 . A Study of the Effectiveness of Language Laboratories :A Preliminary Evaluation in Twenty-one School Systems of the Metropolitan School Study Council . New York :The Institute of Administrative Research ,Teachers College ,Columbia University .
- Kiely ,R . &P . Rea-Dickins . 2005 . *Program Evaluation in Language Education* . New York :Palgrave Macmillan .
- Joint Committee on Standards for Educational Evaluation . 1994 . *The Program Evaluation Standards* . Thousand Oaks ,CA :Sage .
- Mathews ,T . J . &C . M . Hansen . 2004 . Ongoing assessment of a university foreign language program . *Foreign Language Annals* 37 /4 :630-640 .
- Norris ,J . M . 2004 . Validity evaluation in foreign language assessment . Honolulu ,HI :Unpublished Ph .D . dissertation .
- Patton ,M . Q . 1997 . *Utilization focused Evaluation :The New Century Text (3rd ed .)* . Thousand Oaks ,CA :Sage .
- Scherer ,G . A . C . &M . Wertheimer . 1964 . *A Psycholinguistic Experiment in Foreign Language Teaching* . New York :McGraw -Hill .
- Scriven ,M . 1967 . The methodology of evaluation . In R . W . Tyler ,R . M . Gagne &M . Scriven (eds .) . *Perspectives of Curriculum Evaluation* ,pp . 39-83 . Chicago :Rand McNally .
- Smith ,P . D . ,Jr . 1970 . *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction :The Pennsylvania Foreign Language Project* . Philadelphia :The Center for Curriculum Development .
- Watanabe ,Y . ,M . González-Lloret ,Y . Z . Hill &W . Yang . 2007 ,February . Putting useful evaluation into practice . Presented as Second Language Teaching &Curriculum Center (SLTCC) Teaching &Technology Workshop . Honolulu ,HI :University of Hawaii at Manoa .

Yang, W. 2007. Evaluation of a teacher induction program in a U.S. university English language institute : Towards useful evaluation. Honolulu, HI : Unpublished M.A. Scholarly Paper.

Appendix 1 CHN 111 Course Final Evaluation

1. What do you think of the online teacher and the tutor ? Is he /she effective in helping you reach your language learning goal(s) ,in terms of knowledge presentation ,feedback ,quantity and usefulness of the assignment ? Is he /she approachable to solve your language related problems ?
2. What do you think of the material (textbook ,CD ,DVD ,labwork ,etc .) ?
3. What do you think of the online activities ,in terms of user friendliness ,interactiveness ,and usefulness for your language learning ?
4. How effective is this class overall in helping you to reach your language goals ?
5. What are the overall strengths and weaknesses of this course ?
6. What are your suggestions to improve this course ?

Appendix 2 Sample Students Feedback Form

For :XXX Date :4 /23 /06
 Purpose :To help you improve your language learning in this class .

Areas	Comments
Pronunciation in speaking : Sentence pause Sentence intonation Pronunciation Tone	
Language structure in writing : Accuracy Fluency (length) Complexity (usage of conjunction words and adverbial phrases)	
General performance : Preparation Responsiveness online Attendance Motivation Learning strategy	
Overall Comments : Instructor s signature :	

Appendix 3 Final Course Evaluation and Needs Analysis Survey for CHN 112 Spring 2006

This survey will help the instructor to improve the course based on your evaluation and your needs . Your identity is anonymous .

Section One :Evaluation of the course components

I . The online instructor . On a scale of 1 -10 ,1 means strongly disagree ,10 means strongly agree .

Aspect	Rating	Comments (optional)
Example humorous	9	He tries to be serious ,but he s actually very funny
1 Effective		
2 knowledgeable		

Aspect	Rating	Comments (optional)
3	willing to help	
4	responsive	
5	helpful	
6	good teaching style	
7	experienced	

8 . I think the online instructor could be more helpful if he _____

II . The tutor . On a scale of 1 -10 ,1 means strongly disagree ,10 means strongly agree .

Aspect	Rating	Comments (optional)
1	effective	
2	knowledgeable	
3	willing to help	
4	responsive	
5	helpful	
6	good teaching style	
7	experienced	

8 . I think the tutor could be more helpful if she _____

III . The material . On a scale of 1 -10 ,1 means completely useless ,10 means extremely useful .

Aspect	Rating	Comments (optional)
1	the textbook	
2	the textbook CD	
3	the lab workbook	
4	the lab workbook CD	

5 . I think the material could be more improved or made more useful by _____

IV . The online components . On a scale of 1 -10 ,1 means completely useless ,10 means extremely useful .

Aspect	Rating	Comments (optional)
1	the cafe discussion board	
2	the weekly schedule	
3	listening exercise	
4	Q &A forum	
5	tutoring preparation dialogue	
6	follow up tutoring tasks	
7	language clinic tasks	
8	quizzes	
9	recording components	
10	teacher s feedback	

11 . I think the online activities could be improved or made more useful by : _____

V . The tutoring sessions . On a scale of 1 -10 ,1 means completely useless ,10 means extremely useful .

Aspect	Rating	Comments (optional)
1 activities		
2 organization		
3 language instruction		
4 atmosphere		

5 . I think the tutoring session could be improved or made more useful by : _____

VI . Overall . On a scale of 1 -10 ,1 means absolutely no improvement through /in this course ,10 means great improvement through /in this course .

Aspect	Rating	Comments (optional)
1 pronunciation		
2 expressing my own meaning		
3 listening comprehension		
4 language structure knowledge		
5 sense of community		
6 confidence		
7 interest in China and Chinese		

8 . The most motivating components in this class are : _____

9 . The least motivating components /elements in this class are : _____

Section Two :Evaluation of the improvement efforts this semester

1 . What do you think of the extracurricular 15 -minute conversation practice ? (please make the selection bold)

- A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful

2 . How many times do you want to have such practice per semester and when ? (Including zero)

3 . Your feelings about this practice is that : _____

4 . What do you think of the “students feedback sheet ”that you received from both the instructor and the tutor ?

- A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful

5 . How many times do you want to have such feedback per semester and when ? (Including zero)

6 . Your feelings about this feedback are that : _____

7 . What do you think of the summary of students common mistakes at the end of the semester ?

- A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful

8 . Your feelings about this summary are that : _____

9 . What do you think of the uploaded answer keys for the listening lab workbook ?

- A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful

- 10 . Your feelings about this answer key are that : _____
- 11 . What do you think of the practice of writing in paragraphs and producing connected speech ?
 A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful
- 12 . Your feelings about this practice are that : _____
- 13 . What do you think of the casual 5 -minute chat in Chinese at the beginning of every tutoring session ?
 A . Completely useless B . Kind of useless C . Somewhat useful D . Very useful
- 14 . Your feelings about this practice are that : _____
- 15 . What do you think of checking your textbook exercise online instead of at the tutoring session ?
 A . Much less satisfied B . Rather less satisfied C . More satisfied D . Far more satisfied
- 16 . (For those who took CHN 111 last semester) How satisfied are you with this course this semester compared to last semester ?
 A . Much less satisfied B . Rather less satisfied C . More satisfied D . Far more satisfied
- 17 . (For those who didn't take online CHN111 last semester) How satisfied are you with this course this semester compared to Chinese class(es) you took before (skip is not applicable)
 A . Much less satisfied B . Rather less satisfied C . More satisfied D . Far more satisfied
- 18 . What are your recommendations to help the course improve next semester ?

Section Three ;students needs analysis

Textbook topics . The following are the topics we learned this semester . How useful are these topics to you ?On a scale of 1 4 ,1 =completely useless ,2 =kind of useless ,3 =somewhat useful ,and 4 =very useful . Also give comments on why certain unit(s) or subunit(s) are not so useful to you .

Topic	Rating	Comments
EXAMPLE ;Unit 8 ;This is my school	3	For section a ,some vocab are not so useful , like roller pen For section b ,we also want to learn the subject name that we are studying , like Botany ,Arts ,Business etc .
5 Let s Go Shopping		
6 What Do You Like (preferences , hobbies)		
7 This Is My Home (Location , rooms , furnishing ,daily routine)		
8 This Is My School (Classroom Objects , Subjects ,locations ,and persons in school)		
9 Let s Eat In a Restaurant (food name , style ,dish name ,drink ,menu)		
10 Where Shall We go (establishments in a town ,distance)		

For the next year of Chinese learning ,what topics you would like to be taught ?
