5-31-2010

Experiences of Foreign Language Teachers and Students Using a Technology-Mediated Oral Assessment

Jeannie Ducher
University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the American Studies Commons, and the Secondary Education and Teaching Commons

Experiences of Foreign Language Teachers and Students
using a Technology-Mediated Oral Assessment

by

Jeannie Ducher

A thesis submitted in partial fulfillment
of the requirements for the degree of
Education Specialist
Department of Secondary Education
College of Education
University of South Florida

Co-Major Professor: Carine M. Feyten, Ph.D.
Co-Major Professor: Marcela van Olphen, Ph.D.
Linda Evans, Ph.D.
Wei Zhu, Ph.D.

Date of Approval:
March 29, 2010

Acknowledgements

My heartfelt thanks go to my husband Steve for his great fortitude and understanding during the completion of this degree. His love and support were my beacon of hope, as was the unconditional love of my boy, Alex. My gratitude also goes to my friend and major professor, Dr. Carine Feyten, whose forceful friendship and delicate prodding helped keep me on the road to achieving my goal. To the other members of my committee, my thanks for their facilitative help and needling moral support. Last but not least, a hug to my entourage of doctoral friends, as our shared tribulations forged friendships steeped in great humor and professional competence, and were largely instrumental in making obtaining this degree a highly enjoyable experience.

Table of Contents

List of Tables

List of Figures

**Experiences of Foreign Language Teachers and Students
using a Technology-Mediated Oral Assessment**

Jeannie Ducher

ABSTRACT

The development of the speaking skill at the lower levels of proficiency is seldom

assessed as a matter-of-fact in the foreign language classroom, for reasons of

impracticality and difficulty of implementation. Although the practice of the speaking

skill is an important part of current approaches to the teaching of foreign languages,

issues of time and logistics often prohibit the direct evaluation of the skill in a manner

consistent with best practices, which purport that practice and assessment must be closely

aligned, and that students benefit from self-evaluation and teacher feedback. Classroom

research has shown that a skill that is not assessed, although practiced in class, sends the

implicit message that this skill is not as valued as others that are the object of evaluation.

This project presents the rationale, background literature and methodology to use current

computer technologies in an attempt to offset these preventative issues, and to offer

foreign language students and teachers a flexible model to conduct evaluations of

students' oral development in a practical, authentic and valid manner, with opportunities

for constructive feedback and tracking of students' progress.

Chapter 1

Introduction

Omaggio-Hadley (2001) states that

> Many classroom teachers feel that oral exams are among the most difficult
> types of exams to create, schedule, administer, and grade, especially when
> they have large classes or many classes in the course of a day. For these
> reasons, some classroom teachers consider oral tests impractical and do not
> attempt to test oral skills in any regular or consistent fashion. Yet it is clear
> that one of the priorities in the language teaching profession in recent years
> has been the development of oral proficiency. If we hope to convince
> learners via our testing program that communicative language use is a major
> goal of instruction, we will need to administer some type of oral test, if only
> once or twice a semester (pp. 431 & 433).

Statement of the Problem

The concepts of proficiency-oriented foreign language instruction and communicative competence require the development of all facets of communication skills in a foreign language course, including, and maybe foremost, that of speaking ability.

Most language learners will declare the ability to communicate orally with native speakers as their primary motivational goal and expected outcome for taking a language course. Pedagogically, then, it stands to reason that a skill that is highly desirable and highly desired should be practiced, and consequently, assessed. Not assessing a skill practiced in class sends the implicit message that this skill is not as valued as the others that are the object of evaluation.

The assessment of the speaking skill, however, has been notoriously and traditionally overlooked throughout the history of foreign language learning and teaching for reasons of impracticality and difficulty of implementation (Barnwell, 1996; Egan, 1999). Although the practice of the speaking skill may have been an important part or even a primordial component of particular approaches to the teaching of foreign languages (e.g. the audiolingual method), the evaluation of acquired oral skills were and still are seldom attempted directly, as the time and logistics involved are prohibitive in an increasingly tightly-packed foreign language curriculum (Egan, 1999). Even the relatively new focus on authentic assessment has not succeeded in overcoming the challenge to directly evaluate oral achievement in the classroom. Research into best practices suggests that it is not pedagogically sound to attempt measuring oral proficiency through indirect methods such as paper and pencil tests (Flewelling, 2002). As a result, most instructors at the low and intermediate levels are often left with unsatisfactory ways of measuring oral achievement.

Accordingly, before the proficiency movement and the consequent stress on authentic assessment, evaluating oral competence was often indirectly attempted through tasks and activities in the three other skills, reading, writing and listening (Barnwell,

2

1996). The proficiency movement initiated in the 1970s made the ability to functionally

and appropriately communicate in the target language the main goal of language learning,

and brought attention to the need to evaluate directly the oral proficiency resulting from

instruction. This direct evaluation, where speaking ability is assessed through speaking

tasks, has proven to be a challenge.

There is, to this day, no consensus as to a direct achievement tool to assess oral

progress in the foreign language classroom that can withstand the constraints of time and

impracticality. The well-known Oral Proficiency Interview (OPI), developed by the

American Council on the Teaching of Foreign Languages (ACTFL), is indeed a direct

assessment of speaking ability, but it is independent of classroom-specific instructional

goals. Moreover, its 30-minute, two-on-one format precludes it from being a practical

classroom tool. The SOPI, or Simulated Oral Proficiency Interview, was developed on

the same premise as the OPI, but as a more practical tool for classroom use, as it involves

the use of tape recorders to play prompts and of cassette tapes for students to record their

oral answers. It does have the advantage of allowing the assessment of several students at

once, yet it is a measure of oral proficiency, not tied to any particular curriculum, and is

ill-suited to provide information as to progress in oral development, especially since

student production is sent outside the classroom for evaluation. Moreover, it is not meant

to assess the oral proficiency of novice learners of a language, as it has been developed to

assess "the examinee's ability to perform different functions at the ACTFL Intermediate,

Advanced, and Superior levels." (Center for Applied Linguistics, 2006,

www.cal.org/resources/digest/0014 simulated.html).

Issues as to the validity of the current oral proficiency assessments have also been raised. The interview format, in particular, has been decried as not providing the real-life context of a conversation it claims to do; it is an unbalanced exchange of instructor questions/student answers following the strict pattern of survey research interview (Johnson, 2001). As such, it leaves little opportunity to the examinee to initiate topics, or include any other social uses of language as would be found in real-life conversational exchange. Normal conversation, even when in interview format such as a job interview, has for purpose the exchange of voluntary information between interlocutors (Perrett, 1990). The one-on-one interview speech as is being practiced in second/foreign language testing overtly simulates the exchange of information, but, according to Perrett, this is only its secondary purpose; it primarily — and covertly — aims at uncovering student ability to use the language appropriately.

Advances in technology and the now ubiquitous presence of computers in education have prompted the development of several computer-delivered assessments of oral proficiency. The Center for Applied Linguistics (CAL) has been developing the COPI (Computerized Oral Proficiency Instrument) as an alternative to the OPI and the SOPI in Arabic and Spanish. The assessment is based on a computer-adaptive algorithm that determines, according to students' answers to probing, level and above-level questions, the level of competency a student has reached. It is not suited for low levels of proficiency. The STAMP (Standards-Based Measurement of Proficiency), developed by the Center for Applied Second Language Studies (CASLS), is an online assessment of reading, writing, listening, and speaking proficiency for novice low to intermediate high levels of proficiency. Although the test items are based on authentic materials and

4

realistic tasks, it is a summative assessment not tied to a particular curriculum, and cannot be substituted for an achievement test of oral ability. Other computer-delivered assessments of oral proficiency such as the TOEFL and BEST Plus exhibit the same qualities as the above-mentioned instruments, as well as the same limitations for a classroom achievement test.

On the other hand, the classroom-developed oral interview, as it is widely practiced at the secondary and tertiary levels (Flewelling, 2002), is a measure of classroom achievement within a proficiency-oriented approach to language instruction. As a one-on-one, face-to-face oral interaction of approximately 3 to 5 minutes between instructor and student, it is a time consuming process that often requires more than one class period to be completed. As the time constraint makes it difficult to administer more than once or twice a semester, learners come to the oral interview task unprepared and with high anxiety. Indeed, students seldom practice speaking in the classroom in the interview format, as classroom speaking activities tend to be paired dialogues meant to practice specific features of the language in context. There is dissonance between practice and assessment, which runs contrary to current theories about good practice in assessment. Student unfamiliarity with the task combined with the high stakes value of the test renders the assessment unreliable. Moreover, learners and assessor are caught in the "speaking moment", forcing the assessor to judge student performance holistically with little opportunity for constructive feedback. It is possible to try to obtain a more reliable evaluation of student performance by audio or video recording the interaction, but the instructor's investment of time into the assessment is then doubled, adding to the impracticality of the assessment.

More congruent with current best practices in teaching and assessment is the concept of authentic assessment, which occurs in the classroom as students practice a skill. Classroom observations of pairs or small groups of students practicing speaking on communicative tasks directly measure progress against the course curriculum. Students are being evaluated on tasks with which they are familiar, the practice is non-threatening as instructors may observe from afar, and thus the exercise tends to yield better quality of oral production. This type of assessment, however, does not allow for feedback nor does it allow the teacher to revisit the language produced. Moreover, unequal abilities in speaking between students may negatively effect scoring. The teacher usually evaluates the overall quality of the target language in use, as well as the effort expanded by the students to accomplish the task with a checklist. Such a holistic, imprecise assessment is of little use to help learners progress in their speaking ability.

Providing foreign/second language instructors with an authentic assessment that will give them detailed and pertinent information of their students' oral progress in the target language as well as address the issues of practicality and validity is paramount in a proficiency-oriented approach to language instruction. The current developments in multimedia computer technologies may present a viable alternative to the issues preventing oral achievement assessments to be performed on a regular basis.

Purpose of the Study

This project endeavors to explore instructors' and students' experiences with a technology-mediated oral achievement assessment system that allows ongoing practice

and tracking of student progress. It will also investigate the choices made by teachers among the various options offered by the assessment system, such as e-portfolios, oral or written feedback, oral or written prompts, etc. The assessment system will make use of the multimedia capabilities of computers, thus providing students with contextualized situations as they were presented during instruction. As students practice in class and revisit the same speech situational features in a similar format, task unfamiliarity ceases to be an issue. It can be hypothesized that some of the anxiety brought by task unfamiliarity and the interview format will be alleviated, making the assessment of student achievement more valid and reliable. The computerized assessment will be web delivered, giving students the opportunity to access the test in their own time, whenever they please, as part of their unit assignments. Class time will then no longer be set aside to administer the assessment and the teacher is no longer involved in the logistics of administering the test, therefore freeing instructional time. Lastly, as students record their production of the target language in an e-portfolio available to both teachers and students for review, teachers have the opportunity of giving students pertinent feedback on their performances and assess progress over time. Instructors' and students' experiences with the computer-mediated oral assessment, and congruence between best practices in assessment, best practices in instruction, and issues of assessing the oral development of beginning students of French will be analyzed.

Research Questions

Two main questions will guide this inquiry:

Research Question 1: How does technology affect beginner-level foreign language instructors' and students' experiences with oral language assessment?

Research Question 2: How does a computer-mediated oral assessment impact the alignment between the challenges of oral assessment and best practices in assessment?

In order to help answer these two overarching questions, three sub-questions will be answered separately:

1. How do instructors of French and Spanish conceptualize the assessment of speaking achievement?
2. How do instructors of French and Spanish experience a technology-mediated assessment of speaking?
3. How do beginner students of French and Spanish experience a technology-mediated assessment of their speaking development?

Significance of the study

This research is significant for a number of reasons. The debate over the best ways and instruments to assess oral proficiency has been ongoing since the birth of language testing as a field per se (Barnwell, 1996). The notion of oral proficiency testing has been contested for the little information it provides as to classroom achievement. Assessing oral development within the confines of a specific curriculum does not

challenge the concept of proficiency-oriented instruction. Proficiency still remains the ultimate goal of foreign language instruction, but proficiency instruments regain their rightful place and purpose, which is the evaluation of the attainment of a level of functional ability at the end (or the beginning, in the case of placement tests) of a sequence of study, not as an inadequate substitute for classroom achievement tests. Providing an oral achievement instrument fills a gap that is deeply felt in the profession.

Indeed, most foreign language instructors are well aware of the importance and necessity to evaluate not only the speaking skill in their practice, but its development over the course of instruction. Teachers, however, are confronted with the fact that there does not exist an instrument or a model that is easy to implement in terms of time and logistics, that provides useful information in terms of student oral development, and that will allow teachers to give students constructive feedback. With the model of oral achievement testing that this study proposes to research, teachers will be provided with an online test designed for the purpose of oral achievement, with the capability of being easily tailored to their instructional curriculum and needs. The e-portfolio format permits the creation of a progressive profile for student oral development over the entire course of study, enabling teachers to make pertinent instructional decisions through formative assessment. Students will have further opportunity to be involved in their oral skills development by being able to listen to their own speech as they record it, on the spot as well as weeks later, thus self-evaluating their own progress. They will be able to read the teacher's feedback on their performance, be it pronunciation, lexical, grammatical, syntactic or pragmatic issues related to the functionality of speech. This project therefore

provides an opportunity for teachers and students to make more of their instructional and learning time.

The format of the oral achievement test is familiar to teachers who have attempted a direct evaluation of their students' oral skills either through interviews, observations, role-playing, or in a language lab. The computer-mediated oral achievement assessment functions within two of the three modes in which students must become competent according the ACTFL Proficiency Guidelines–Speaking (1999): tasks require that student interact with the language in the presentational mode and the interpretive mode. In these two modes, the computer can play the role of benevolent, non-threatening agent as questions are recorded in the target language for students to listen to and respond to. As live interaction at the lower levels of proficiency between assessor and students tends to intimidate the interviewee and result in inauthentic interaction and anxiety-laden oral production, limited interaction through the computer where students control some of that interaction (e.g. listening to the question several times), may help assuage some fears associated with speaking and testing. Tasks in the two modes may include giving directions from a map, introducing oneself at a party, or ordering food at a restaurant. Students will have practiced these very tasks in class, a practice that is congruent with best practices in assessment as well as in language instruction.

The familiarity of most instructors with the format of the oral achievement test has another implication: it may be less threatening to instructors than other evaluation models that have been offered in the past to assess the speaking skill. Pair or small group evaluations, classroom observations, lab testing, role playing may be a departure from some teachers' realm of experiences, as many current foreign language practitioners went

through their language studies without being assessed in their oral language skills. Given the limited amount of time dedicated to professional development and the necessary survival skills that teachers need to develop in a profession where time is forever being more compressed to increase the amount of instruction, there is little time for teachers to try some new instructional practice and/or assessment that is widely different than their developed practice. Trying to assess oral achievement in a manner that is a great departure form their realm of experience may be to open themselves to feelings of insecurity, uncertainty and frustration, notions with which most teachers are uncomfortable. Therefore if we want to develop the practice of assessing the speaking skill in the classroom, it is important that the instrument be practical as well familiar to practitioners.

Lastly, this research is significant in that it may bring some insight into the washback effect of assessment onto instruction (Gates, 1995). Few studies have investigated the effects of assessment on instructional practices, and none have looked at oral achievement assessment and feedback and its effect on foreign language teaching and learning. It will be interesting to see how teachers use the information they gather from the ongoing assessment of their students' oral development to inform their practice, as well as whether and how students benefit from the availability of feedback.

Limitations and Delimitations of the Study

The participants of this research study will be mainly instructors of French 1 and French 2 in a large urban university whose interests lay in furthering their instructional

knowledge as well as obtaining a better oral assessment tool. Selected students will be

asked to participate in the study, and all will perform their oral assignments as part of

their regular curriculum. However, most of the data collected will be through interviews,

and only interested students will participate, given a small incentive. The findings of this

study will therefore not be generalizable to students and teachers other than beyond those

involved in the research. Inferences, however, will be made about learners and instructors

at the same level of instruction and proficiency who might benefit from the insights

gained through the study.

Zhao (2003) in his meta-analysis of the recent developments in technology and

language learning, argues that most of the research on technology in language education

focuses on the college level, leaving the K-12 levels unrepresented. Issues of access

prevent this study to be conducted at the K-12 level, but it is endeavored to replicate this

study in the middle/high school levels as further research.

Definition of Terms

Assessment   An ongoing process aimed at understanding, documenting

and improving learning (Stiehl, 2002). A combination of

assessment techniques designed to provide a

comprehensive and useful picture of student achievement.

Achievement Assessment	Also termed classroom assessment. A series of formative tests aimed at measuring a learner's mastery of material tied to a specific curriculum, and guided by the professional judgment and knowledge of the teacher (Angelo & Cross, 1993).
Authentic Assessment	A form of assessment in which learners use their knowledge and skills to perform in meaningful, real-life tasks and activities. Usually evaluated with the use of a rubric.
Direct Assessment	A form of assessment which requires students to perform the skills involved in determining their ability in the particular skill. Requiring students to produce speech is a form of direct assessment of speaking ability.
Formative Assessment	A process of assessment techniques aimed at giving teacher and learner a progressive view of student development over a period of time. The information gained is used by both teacher and student to effect change in teaching and learning.
Indirect Assessment	A form of assessment in which examiners or raters infer student abilities, knowledge or values rather than observe direct evidence. Assessing students' oral ability by

administering a listening comprehension test or a multiple-choice test is an example of an indirect assessment.

Oral Prochievement Assessment: an oral examination that aims to evaluate the development of oral proficiency from the course content and objectives.

Proficiency Assessment   A summative test aimed a determining a learner's attainment of proficiency in the target language. These assessments are not tied to any particular curriculum, but are usually indexed on the *ACTFL Proficiency Guidelines*.

Summative Assessment   A comprehensive form of assessment occurring at the end of a course or program of study. It provides accountability and checks students' level of attainment in learning.

Washback   "The effect of testing on teaching and learning" (Gates, 1995, p.101).

Chapter 2

Literature Review

Introduction

The evaluation of the speaking skill in the foreign language classroom is a challenging endeavor for both instructors and students, mainly for reasons of logistics and practicality. The current computer technologies available to classroom practitioners present several advantages that may provide a satisfactory alternative for teachers to check their students' progress in speaking ability. This chapter's discussion presents the past and current literature pertaining to the evaluation of the speaking skill within a proficiency-oriented approach to language learning and teaching.

Chapter 2 is organized in four major sections. The first section presents the proficiency-oriented paradigm within which most foreign language instructors teach, as well as the proficiency movement's underlying assumptions and concept of communicative competence. The second describes the best practices and strategies that lead to effective communication skills in the target language. The third section discusses the various issues and challenges pertaining to the assessment of the speaking skill. It describes current practices in oral assessments and the instruments available to

practitioners, as well as argues for the necessity to adhere to principles of best practices in assessment. The fourth and final section interprets findings of the CALL (computer-assisted language learning) and CALT (computer-assisted language testing) literatures as they relate to the assessment of the speaking skill. It also presents the different computer-mediated oral assessments available and discusses the advantages and disadvantages of each.

The Speaking Skill within Language teaching and learning

There are two branches to the current research in language learning; Shrum and Glisan (2010) separate them into two categories. They call the experimental and classroom-based research that has focused on the acquisition of language by individual learners "acquisition as a *cognitive process* that occurs in the individuals' brains (Chomsky, 1968; Corder, 1973)" (Shrum & Glisan, 2010, p.12). In contrast, they call the more recent research steeped in Socio-Cultural Theory "acquisition as a *social process* that occurs during interaction with others (Firth & Wagner, 1997, 2007; Hall, 1997; Swain & Deters, 2007)" (p.12) as this research is conducted within and outside of classroom settings. This study will be looking at language learning both as a cognitive process, or individual achievement and as a collaborative achievement within a community of learners, as both theoretical branches are aligned with the proficiency-oriented approach that is most currently used in the foreign language classroom.

Within this approach, the emphasis in learning a foreign language is placed on contextualizing instruction. Language is used purposefully to effectively convey meaning

16

in authentic communicative acts. As a conceptual framework, this approach carries a number of assumptions that will be discussed in details below.

The proficiency movement came to be as a radical shift in the conceptualization of language studies. Until the late 1960s, the goals and outcomes of mainstream language learning programs were tied to the achievement testing of material covered in class. The prevalent thought until this point in language teaching had been to provide students with the tools they needed to analyze and understand the functioning of the target language (TL). Mastery of the language was then thought to be obtained from the metacognitive understanding of grammatical rules of the language, which, if one applied himself or herself enough, would lead to functional proficiency. The launch of the Russian Sputnik in 1968 brought in the realization that this method of teaching a foreign language did not result in the desired proficiency outcome. The reaction to Sputnik brought an end to the faculty psychology approach to language teaching. The 1960s were dominated by the audiolingual method, or ALM, originating in behavioral psychology. Emphasis was placed almost exclusively on aural/oral skills drills, which were developed through rote memorization of set sentences in determined situations. The following decade was a rich period of intense development in the field, influenced by the fields of cognitive psychology, generative grammar, and semantics; a number of approaches to language instruction arose and were tried out, with varied and mixed results. The Council of Europe's 1977 innovative program design and new methods for language instruction offered new perspectives and vision for language teaching. In the midst of these effervescent times the drive to unify and promote excellence in education jumpstarted the movement to bring standards to language teaching. Finally two reports, *Strength through*

Wisdom from the President's Commission on Foreign Language and International Studies and the Tongue Tied American authored by Senator Paul Simon were both released in 1979, shocking the profession with their descriptions of the state of American second and foreign language instruction. The nation then realized that a drastic shift was needed from what was being taught to what students could do with the language as they exited a language program in order to enact significant changes into the effectiveness of language instruction. The proficiency-oriented movement was born. Omaggio Hadley's table below (Table 1) presents its main assumptions under the form of five hypotheses.

As can be seen in Table 1, the principal underlying assumption of the movement is the direct application of coursework learning to life-like situations. Course objectives, activities, tasks and assessments are designed to apply linguistic and paralinguistic features of the studied language to real-world situations and contexts. Learning, then, ceases to be measured in terms of number of credit hours, semesters or years of study, but is realized through practical, life-like communicative skills (Schultz, 1986). Becoming orally proficient is tantamount within this approach to language teaching and learning, as can attest the numerous efforts the field has deployed over the last three decades to devise, develop and refine oral proficiency guidelines and oral proficiency assessments.

Table 1. Orienting Instruction Toward Proficiency, from *Teaching Language in Context* by Omaggio Hadley, A. (2001).

Hypothesis 1	Opportunities must be provided for students to practice using a range of contexts likely to be encountered in the target culture Corollary 1. Students should be encouraged to express their own meaning as early as possible after productive skills have been introduced in the course of instruction. Corollary 2. Opportunities must be provided for active communicative interaction among students. Corollary 3. Creative language practice (as opposed to exclusively manipulative or convergent practice) must be encouraged in the proficiency-oriented classroom. Corollary 4. Authentic language should be used in instruction wherever possible.
Hypothesis 2	Opportunities should be provided for students to practice carrying out a range of functions (tasks) likely to be necessary in dealing with others in the target culture.
Hypothesis 3	The development of accuracy should be encouraged in proficiency-oriented instruction. As learners produce language, various forms of instruction and evaluative feedback can be useful in facilitating the progression of their skills toward more precise and coherent language use.
Hypothesis 4	Instruction should be responsive to the affective as well as the cognitive needs of students, and their different personalities, preferences, and learning styles should be taken into account.
Hypothesis 5	Cultural understanding should be promoted in various ways so that students are sensitive to other cultures and are prepared to live more harmoniously in the target-language community.

The common metric determining oral communicative proficiency is the ACTFL

Oral Proficiency Guidelines—Speaking, devised in 1986 and subsequently revised in

1999. These guidelines describe in operational terms what students are able to do with the

target language at different levels of proficiency. These levels were experientially established according to a natural progression within the formal education system. The *Guidelines* have been represented as an inverted pyramid to help educators conceptualize the unequal amount of knowledge necessary to go from one level to the other. Students, at the lower levels of proficiency, will quickly progress from Novice-Low to Novice-Mid to Novice-High, as a result of the first two semesters of college studies (Magnan, 1986). It will take exponentially more time and breadth in their knowledge of the target language to move to the subsequent levels of proficiency, the more challenging being the last level of proficiency, or Advanced-High (Breiner-Sanders et al, 2000).

One of the stated goals of proficiency-oriented language learning programs, and one of the most widely held expectation for language learners, is the achievement of a certain level of oral proficiency by the end of a course of study (Omaggio-Hadley, 2001). Unrealistic expectations for achievement of proficiency have led many students and programs to doubt the effectiveness of language learning programs in the past, but new understandings in second language acquisition as well as the descriptive nature of the ACTFL *proficiency Guidelines—Speaking* (1999) allow instructors and students alike to set attainable goals for language proficiency depending on overall course objectives, intensity of the language learning experience, and course format and methodology.

Advanced

Advanced-Low

Intermediate-High

Intermediate-Mid

Intermediate-Low

Novice-High

Novice-Mid

Novice-Low

Novice-High: Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements. Can ask questions or make statements involving learned material. Shows signs of spontaneity although this falls short of real autonomy of expression. Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. Pronunciation may still be strongly influenced by first language. Errors are frequent and, in spite of repetition, some Novice-High speakers will have difficulty being understood even by sympathetic interlocutors.

Novice-Mid: Oral production continues to consist to isolated words and learned phrases within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words. Speaker may have some difficulty producing even the simplest utterances; Some Novice-Mid speakers will be understood only with great difficulty.

Novice-Low: Oral production consists of isolated words and perhaps few high-frequency phrases. Essentially no functional communicative ability.
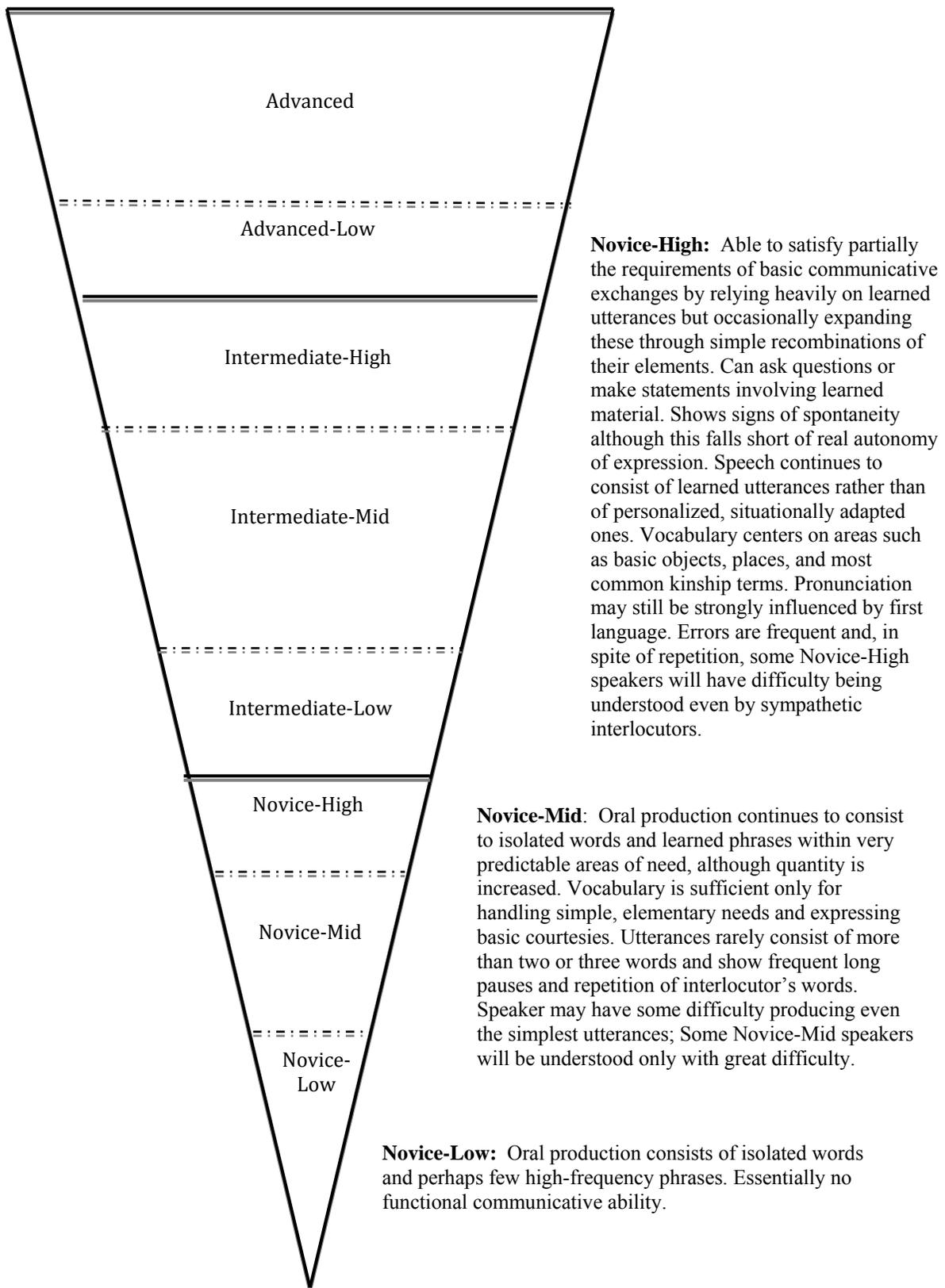
Figure 1. Inverted Pyramid, ACTFL Proficiency Standards

Learner factors, such as personal goals, motivation, commitment, learning strategies, amount of prior language study and language ability, also factor in the realization of this set of proficiency objectives (Magnan, 1986). It is important to keep in mind, however, cautions Omaggio-Hadley (2001), that proficiency is focused on measurement, and is not a method for language teaching. It does not represent a particular theoretical, philosophical, or methodological approach. As such the ACTFL *Guidelines* (1999) do not prescribe a particular methodology to language teaching; they do, however, directly influence the methods and procedures that are used in the classroom to achieve the goals of proficiency-oriented instruction.

These methods and procedures have been derived from the research done on the speaking skill and language acquisition, in particular in the domains of interaction, negotiation of meaning, and learning strategies. Speaking is the interactive construction of meaning in which information is not only produced but also received and processed (Brown, 1994; Burns & Joyce, 1997). Speaking is context-dependent, in that its form and meaning are determined by the purposes for and setting of the interaction, as well as by the personalities and prior experiences of the participants to the speech act. As such, successful speech requires the development of linguistic competence, which includes grammatical knowledge, pronunciation and vocabulary, as well as sociolinguistic competence, or knowledge of when, why, and how to generate speech. Two other competences, strategic competence and discourse competence, were identified by Canale and Swain (1980) as making communicative competence, a concept proposed by Hymes in 1972. This concept links a person's competence to use the language effectively not only to the knowledge of grammar and lexicon but also to the knowledge of the social

conditions in which interaction takes place, such as setting, participants, and goals for the communication act. Strategic competence regards the knowledge of particular strategies to repair or compensate for breakdowns in communication, and includes the use of both verbal and non verbal strategies, such as circumlocution, rephrasing, etc. Discourse competence concerns the ability to formulate coherent, articulated thoughts and expressing them in cohesive form. This concept of communicative competence is central to the proficiency-oriented approach to instruction of languages. It entails the notion of a dynamic interaction between interlocutors in which information is exchanged through appropriate linguistic and paralinguistic input (Savignon, 1972). It emphasizes the interrelationship between speakers, context, intent and meaning, as communication is intrinsically negotiative in nature (Savignon).

The speaking skill has been seldom investigated, as Crookes (1991) remarked: "the role of output (i.e. production or use) in the development of second language proficiency has largely been ignored or denied" (p.117). Most of the research done about the role of output in language acquisition and in developing proficiency occurred in the areas of negotiation of meaning and the comprehensible input-output hypotheses. To Krashen's 1982 statement that "Acquisition occurs, according to the input hypothesis, when acquirers understand input for meaning, not when they produce output and focus on form" (p.117), Long (1983), among others, opposed the proposition that the opportunity to negotiate communication issues facilitates language acquisition. Swain's comprehensible output hypothesis (1985), emerging from her research on oral production skills with immersion students, posited that the development of a learner's communicative competence depends of both comprehensible input as well as

comprehensible output. Swain went a step further when she contended that forcing

learners to produce comprehensible output as a result of the interaction between

interlocutors also facilitates language acquisition. She insisted on the importance of

requiring learners to produce the target language for language acquisition:

> "Simply getting one's message across can and does occur with
>
> grammatically deviant forms and sociolinguistically inappropriate
>
> language. Negotiating meaning needs to incorporate the notion of being
>
> pushed toward the delivery of a message that is not only conveyed, but
>
> that is conveyed precisely, coherently, and appropriately. Being 'pushed'
>
> in output, it seems to me, is a concept parallel to that of the $i + 1$
>
> comprehensible input" (pp. 248-249).

Swain adds: "…using the language as opposed to simply comprehending the

language may force the learner to move from semantic processing to syntactic

processing" (p.168). The concept of comprehensible, pushed output is significant in the

literature on learning strategies, which found that learners need to consciously attend to

both the semantic and syntactic processes to produce output as they strive to understand

input (Rubin, 1987; Chamot, 1987; Dickinson, 1987; Holec, 1985). Ellis (1992)

commented that "'comprehensible input' is not really the result of the *separate*

contributions of the native speaker and the learner but of their *joint* endeavors. The

speech addressed to learners is the result of an ongoing interaction between learner and

native speaker. In this process the interlocutors collaborate in establishing and

maintaining a topic. This has been referred to as *negotiation* (e.g. Tarone, 1981)." (p. 19).

Comprehensible input and comprehensible output then work together to help learners

develop their communicative skills. Breaks or gaps in communication, particularly in trying to produce language, may be useful in triggering cognitive processes that help learn a second or target language (Swain, 1995).

Native or near-native speakers of a language use a number of communication strategies to successfully implement oral communication. The same strategies are available to learners of a foreign language, and must be taught for them to reach strategic competence. In order to progress from incomprehensible output (not precise, coherent, or appropriate) to comprehensible output (Liming, 1990), foreign language learners must learn to use these identified communication strategies, so as to overcome communication problems due to and with limited target language resources. It is therefore essential to teach students at the very beginning of their language studies appropriate communication strategies, to compensate for their initial lack of linguistic competence in the language and enable them to adequately progress in the development of their speaking ability.

The Teaching of Speaking

Speaking in the second/foreign language classroom has enjoyed significantly less interest and research than its counterparts in listening, reading and writing skills (McCarthy & O'Keeffe, 2004). Research into teaching effectiveness has resulted in principles for best practice. It has been found that providing effective practice time in class is essential to the development of oral communicative skills toward proficiency. Payne and Ross (2005) suggest that speaking ability improves when speaking is practiced

in diverse situational contexts, and when the range of topics used fulfills a variety of socio-pragmatic requirements.

Yet upon looking at classroom discourse and interaction, it has been found that the traditional model of teacher-centered instruction, in which the instructor is the purveyor of knowledge and students the recipients of that knowledge, lacks meaningful interaction. Indeed, the prevailing type of teacher-student interaction in teacher-centered classrooms is IRF, or Initiation-Response-Feedback. In this type of interaction, the teacher initiates a question, directed either at a particular student or at any student. The selected student responds, and the teacher gives feedback. Research findings in classroom discourse have shown that instead of giving students opportunities to express themselves in the target language, the model itself exerts some constraints on the interaction, rendering it unsuited to authentic communication. Teacher initiations are usually closed questions (i.e. what color is your pencil?), student response is usually one or a few expected words (i.e. yellow or my pencil is yellow), and teacher feedback often consists of a low evaluative comment as to student performance (i.e. Good, or it is yellow, yes) (abd-Kadir & Hardman, 2007, among others). In order to provide students with more authentic and effective opportunities to develop their speaking skills, the model of classroom discourse and interaction must change to one that is student-centered. A student-centered classroom views students as active agents in charge of their learning, and the teacher as a facilitator of learning, creating the circumstances most propitious to speaking. Such circumstances are tasks and activities that are in essence authentic: they create a context and place students in situations that closely resemble real-life conditions; they allow students to use the whole range of communicative competence available to

26

them at their level of proficiency. At the lower levels of proficiency, students practice the vocabulary and structures in pre-communicative activities, in which student interactions are highly structured yet do place them in real-life contexts. As students' proficiency increase, they practice their skills in more authentic, communicative activities, which allow for more creativity in output. As effective oral skills instruction is concerned with providing learners with the opportunity to practice language in "a range of contexts likely to be encountered in the target culture" (Omaggio-Hadley, 2001, p.179), speaking tasks and activities requiring students to access their higher order thinking skills are devised. Omaggio-Hadley advises to "occasionally introduce material from slightly higher proficiency range so that students can become familiar with it (perhaps for partial or conceptual control), thus preparing themselves for future progress along the scale" (p.183). The contextualization of tasks and activities is an essential tenet of proficiency instruction, as it establishes and reinforces the connection between form and meaning.

Assessing the Speaking Skill

There are different, authentic manners in which oral development may be assessed in the foreign language classroom. It is important here to differentiate between the concepts of proficiency assessment and achievement assessment. Proficiency assessment are essentially summative evaluations, aimed a determining a learner's level of attainment of proficiency in the target language. They are not tied to a particular curriculum, but are usually indexed on the *ACTFL Proficiency Guidelines*. They are usually rather long and exhaustive. The Oral Proficiency Interview (OPI) is a

characteristic example of proficiency assessment, as is the TOEFL (Test of English as a Foreign Language). In contrast, an achievement assessment is closely bound to a course curriculum, and is administered often to measure progress and attainment in the learning of the particular features of language.

According to the ACTFL *Proficiency Guidelines—Speaking* (1999), and in reference to the inverted pyramid (Figure 1, p.19), intending to assess student proficiencies at the novice levels will yield little information as to their overall proficiency, and can only be detrimental to their self-confidence. Testing students' oral development is better served, at this level, with assessments that focus on the articulation of the TL (pronunciation, stress, and intonation), as well as accurate syntactic and grammatical use of selected features. As it is essential, though, to retain the communicative goal of all learning and exchanges in the TL, oral tests are developed as prochievement assessments, a word coined in 1989 by Gonzales Pino, and representing an oral assessment that aims to evaluate the oral proficiency attained directly from the course contents and objectives. A prochievement test is therefore closely aligned to the class curriculum as it unfolds, and is to be administered concurrently with other achievement tests in the FL classroom.

The following is a sample list of oral assessments occurring in the foreign language classroom, which vary in their levels of authenticity. They have been ordered from least authentic to most approximating real-life situations:

- Short answers: students are given visual prompts and must translate what they see into sentences. This type of activity tests grammar and vocabulary accuracy, is quick and easy to grade, but is low on the authenticity scale.

- Telling stories while being videotaped: students must tell a story constructed visually by the teacher. Students have two minutes to familiarize themselves with the pictures before starting their narrations. This type of assessment allows differed grading, each student is graded individually, both in the quality of their oral production as well as in their ability to accurately convey the meaning of the pictorial tale.

- Interviews are assessments engaging at least two participants, usually a teacher and a learner, or a native/near-native speaker and non-native speaker, in an exchange where the assessor asks questions of the person assessed. The protocol of questions can be highly structured, or it can be semi-structured, allowing the assessor some liberty in choosing topics and changes in topics. This type of assessment was thought to have a better chance at eliciting longer and more informative responses from students of a lower level of proficiency in the target language (Blaz, 2000). Interviews tend to tests oral comprehension as well as speech production. When not supported by visuals, gestures, etc, this type of assessment relies heavily on student's listening comprehension skills, and therefore privileges students whose learning styles are predominantly auditory. Interviews are usually given once or twice a semester, as it is impractical to try doing so regularly, and so is summative in essence. Because ACTFL made the OPI its notorious trademark for oral proficiency, and despite issues with validity

and reliability, classroom interviews have taken prevalence over other tasks in the measurement of speaking competence, and represent the type of oral assessment most commonly found in the FL classroom (Barnwell, 1996).

- Listening in on an assigned activity during class practice: students work in small groups on an assigned activity, the teacher listens in, and assess students' oral behavior. This type of assessment usually does not look at special linguistic or paralinguistic features; it considers the overall quality of the target language (TL) produced in a continuum, going from comprehensible language to nonsense, from no English to no TL, and looks at student willingness to complete the assigned task and the effort they make in speaking in the TL. Such assignments are difficult to grade with any consistency. They usually are used for formative evaluation of student progress, and come in complement of other assessments.

- Telling stories (TPRS): where students try to reconstruct in their own words (or with the same words) a story that has been previously narrated and deconstructed by the teacher. Students can add a twist to the story to bring a creative element to the story. These evaluations are formative as they are part of a specific TPRS curriculum.

- Conversations/mini skit with prompt cards: paired or small groups of students are given cards with matching scenarios and must conduct a conversation according to the roles defined on their respective cards. This type of assessment is time consuming. Unequal ability pairing can be a problem for grading reliability, as stronger students tend to dominate the interaction. This can be overcome if the more able students are asked to expand on their side of the conversation (e.g.

rephrase, initiate a change in topic, etc.) when their input in the conversation is met by incomprehension or reaches a dead end. Hughes (1989), however, argues that this it is challenging for the grader to focus equally on both students at once.

- Phone message: students are required to leave a phone message on the teacher's voicemail on a pre-established topic. This differed assignment technique affords teachers the convenience of grading outside of class time, and gives them the possibility of listening more than once to the student production for better reliability in scoring.

- Language lab testing: in such a setting, it is possible to record the oral production of a large number of students at once. Through headsets at individual stations students listen to a set of prerecorded questions and record their answers. The grading is differed, which gives the instructor the possibility to listen to the oral production more than once for better reliability in scoring. There is, however, no interaction with this type of assessment between instructor and student, and no possibility for the instructor to rephrase, prompt, or provide feedback. Technical malfunctions are also a common occurrence, resulting in frustration for the student or the instructor if the recordings are not audible.

Each testing task and situation mentioned above brings different issues with it. These issues will be discussed below.

As was mentioned earlier, developing the ability to communicate orally is a pivotal feature of the proficiency-oriented approach to language instruction. Much time is devoted in foreign language classrooms to placing students in contextualized, communicative situations where they can practice and experience the spoken language in

carefully designed tasks. Yet, the skill is seldom assessed, for numerous reasons. A skill

that is not assessed carries the message that it is not an important feature of instruction,

and has negative implications in terms of motivation and achievement for the students.

Although speaking proficiency might be the one foremost desired outcomes of language

studies, as it is not given the same evaluative attention as the other skills, the practice of

communicative oral skills in the classroom assumes a less serious, less important position

in the learner's mind, and as such does not push students to experiment with the oral

language as much as it does with the written language. It is therefore important to take a

closer look at the different issues contributing to making the speaking skill the most

desired skill and yet the least assessed of all the foreign language skills.

There are difficulties inherent to the nature of the oral skill that make it more

challenging to assess in the classroom than the other skills (Flewelling, 2002). Time and

logistics, as the list of assessments above show clearly, are the most prominent

constraints that have so far prevented the assessment of the oral skill to more than a once

or twice ordeal over a semester or year of FL studies. Whether the purpose of assessment

is to obtain an overall measure of communicative ability in particular tasks or situations,

or determined speech features (grammatical features, vocabulary use, pronunciation,

etc.), such an evaluation needs to happen at the individual or small group level, as more

than three people in a group would not allow the instructor to adequately pay attention to

each student's ability according to his/her part in the task. In testing tasks of two to four

minutes, whether interviews or role plays, to cite only the most common, it takes more

than two full time periods to evaluate an average class of thirty. At the middle and high

school levels, to this time deterrent can be added the logistical problem of keeping the

students who are not being tested learning and behaving. Teachers must stay close to their classrooms and attend to both the oral test and the class, or must ask for another teacher or administrator to assume the responsibility of the class while they conduct the oral test. At the college level, time is often set aside from classroom time and students line up in office corridors.

These time and logistical constraints, in turn, compound the anxiety-producing nature of the skill and its evaluation. There are three aspects of the speaking skill that have the potential to generate anxiety. First and foremost, the very nature of the act of speaking in a different language than one's L1, trying to communicate ideas in a language partially known, is destabilizing to the point of paralysis for some speakers, and can in itself cause debilitating anxiety. Secondly, speaking in a foreign language can be likened to public speaking, as students are required to speak in front of peers and/or instructor(s): public speaking is widely recognized as inducing anxiety. Lastly, being evaluated is anxiety raising in itself. Anxiety and its negative effects on second/foreign language learning have been extensively studied (Brown, 1973; Horwitz et al, 1986; Tobias, 1986; to cite a few). It therefore appears essential to find ways to lower learner anxiety to obtain a sample of speech that is valid and reliable.

When to these affective factors are added unfamiliarity with the task and assessment format, as is often the case, and the high stake nature of the oral assessment due to its rarity in occurrence, little faith can be placed into the validity and reliability of the assessment outcomes. Best practices in assessment suggest that the assessment format be practiced often so as to remove the surprise factor from the evaluation. If students are to be evaluated through a classroom oral interview for instance, they should practice the

oral skill by interviewing each other or being interviewed by the teacher as a matter-of-course. Likewise, task/topic familiarity plays an important role in ensuring that learners are assessed fairly and reliably. Being placed in a testing environment where students are expected to improvise answers to an issue when they have practiced their speaking skill on describing pictures, for instance, is adding a confounding variable to the testing situation. In other words, practice and assessment must be consistently aligned in respect to the task as well as to the test format. Yet, to complete the circular argument, practicing and testing often, with assessments that are both impractical and time-consuming, mean that conscientious teachers must either dedicate an inordinate amount of time to the practice and assessment of the oral skill, or must compromise with the authenticity and validity of the assessment.

Technology has been used in the past to try to offset some of the above-mentioned issues. Students were asked to record their speech on cassette tapes, which were sent subsequently to a testing center where professionals listened and scored their speech. This is currently and commonly being done for Advanced Placement exams for example, or for teacher certification exams. The SOPI (Simulated Oral Proficiency Interview) is another example of such technology-mediated oral examination. Students would record their oral answers to pre-recorded questions to a cassette-tape recorder; student tapes were subsequently sent to a grading center where trained raters evaluated students' quality of utterances. This process proved to be quite onerous, however, and the technology not adapted to relieve issues of practicality in the FL classroom. Newer technologies, however, have come to play a prominent role in education. Could computer technology help with the assessment of speaking?

Ever since the birth of computer technologies, high expectations have been placed on technology to help enhance language learning (Salaberry, 2001). Computers have been particularly useful in traditional drill-and-skill computer-aided instruction (CAI), and are widely used to help with the reading, listening, and writing skills. As technology evolved, becoming more efficient, affordable and available, educators have looked more and more toward technology to improve classroom teaching with materials and experiences that are engaging, authentic and comprehensible. Multimedia computing, the Internet and the web, and speech synthesis and recognition are the current trends in educational technology as 'intelligent' technology models (Zhao, 2003). These models are promising, yet as Zhao points out,

> First, […] these technologies vary a great deal in their capacity, interface, and
> accessibility. […] Second, the effects of any technology on learning outcomes lie
> in its uses. A specific technology may hold great educational potential, but until
> it is used properly, it may not have any positive impact at all on learning. Thus
> assessing the effectiveness of technology is in reality assessing the effectiveness
> of its uses rather than the technology itself. […]Third, […] the effectiveness of
> an educational approach is highly mediated by many other variables—the
> learner, the task, the instructional setting, and of course, the assessment tool.
> Thus, even the same use of a particular technology in a different instructional
> setting may result in different learning outcomes." (p.8).

The issue of effectiveness of technology use in language learning has led to many a disappointment, many of which can be attributed to improper use of the technology. For

technology to be effective, objectives and resources need to be closely matched, its purpose clear, and its implementation rigorous. Zhao (2003), in his review and meta-analysis of recent developments in technology in regards to language learning, has found that "techno-based language instruction can be as effective a teacher-delivered instruction" (p.20), and has the potential, when used appropriately, to positively effect language learning. Chapelle (2001) agrees, stating that technology can and should be used in language learning.

She also advocates for the use of technology, and particularly computers, in language testing. The assessment of language learning through technology, however, is rarer in the foreign/second language classroom. Computerized assessments are generally devised and implemented on a large scale for high-stakes testing purposes, the TOEFL (Test of English as a Foreign Language) or the FCAT (Florida Comprehensive Assessment test) being two of the most notorious examples. Tests of speaking proficiency, aiming at evaluating candidates' overall speaking ability in the target language, have also been developed to be computer-delivered. The COPI, Computerized Oral Proficiency Instrument created by the Center for Applied Linguistics, is being developed and piloted around the country. It aims at emulating the results obtained through the ACTFL Oral Proficiency Interview (OPI) in establishing a candidate's level of proficiency, while doing so in an entirely automated way. All three mentioned high stakes assessments are computer adaptive tests (CATs), derived from the psychometric theory item-response theory, or IRT. CATs are based on the evaluation of item difficulty (an IRT item statistic); they seek to assess a candidate's highest level of competence by establishing his/her knowledge threshold through questions varying in difficulty.

Complex algorithms present a test-taker with questions, and multiple-choice answers, of increasing difficulty in each domain of competence, until the candidate's correct responses can no longer be attributed to chance.

Computer adaptive testing is based on item response theory (IRT), in which test developers use item statistics available from the administration of one test, usually a paper-and-pencil test, to help select similar items in the development and administration of subsequent tests. IRT was the driving force behind computer-assisted test delivery, and as such was highly instrumental in perpetuating the "paper and pencil technology" (Hunt, 1987), or conceptualization of testing revolving around a fixed set of alternatives in answer to a large number of short questions, the multiple-choice test (Chapelle, 2001). The validity of multiple-choice assessments, whether paper-and pencil or computer-delivered, as a sole means of evaluation has been brought into question by numerous test-users and researchers over the years. The issue of construct validity, or the ability to derive trustworthy inferences and uses from the test results, is particularly salient in a model where low as well as high stakes decisions are made daily on the basis of a sole type of assessment, as "[the] use of a single test method can result in systematic distortion of what the test is intended to measure" (Chapelle, p.39). Another issue that the multiple-choice model raises is the effect that item selection has on the materials studied in class. In effect, it can be argued that the washback effect of such assessment instruments is that instruction narrows down to the particular competencies needed to be demonstrated on the test, to the detriment of other pertinent domains. Some (e.g. Linn, Baker, & Dunbar, 1991) have decried such an effect on the measurement of language, which is cognitively complex. As a result, Nitko (1989) suggested to mesh instructional practices with their

assessment, and Chapelle (2001) stated that "[…] when language testing leaves behind the well-established technologies of the multiple-choice test, it must develop new test theory (e.g. Mislevy, 1993a; 1994) which must rely heavily on theory and practice in applied linguistics" (p.40).

Criteria for Evaluating Technology-mediated Language Assessment

Computer psychometric models do not lend themselves easily to the measurement of the language productive skills. Use of the language in terms of appropriateness and correctness for purposes of communication cannot readily be measured through a computer model because of the inherent creativity and multidimensionality of language, confronted with the current limitations of the computer models and capacities. Other assessment models must be envisioned and devised, attending to the holistic characteristics and qualities of language concurrent with meaning-based approaches to language learning and teaching. Chapelle advocates for creativity in devising new ways to assess language development, and supports the establishment of guiding principles or criteria for the design of assessments "that can be considered positive qualities of a test as well as methods for their evaluation" (pp.42-43). Bachman and Palmer (1996) have determined six types of arguments as guiding principles for the development and evaluation of language assessments. In lieu of using test validity and validation as the underlying principle for test evaluation, they argue that test usefulness is a more operationable and useful construct for test evaluation. The criteria for test usefulness are

stated in Table 1 (Bachman and Palmer, 1996) and will be used as guiding principles for the current study.

Although ideally all six criteria of test usefulness would be equally emphasized, it is important to note that in reality some criteria will assume more importance than others, depending on the purpose of the test (Chapelle, 2001). As each participant in the study, the language instructors, may have different intentions and purpose for the prochievement tests, their test choices and resulting inferences as to the usefulness of the test will be indicative of the relative importance of each criteria. Bachman and Palmer's six criteria will therefore provide the overarching framework for the interpretation of data in the study.

Moreover, the method of assessment, or the medium through which a test is administered to examinees, is an important dimension of testing. According to Bachman (1990) and others, test method directly effects examinees' performance. Two assessments, in appearance similar, could be measuring different abilities in examinees because of the test method effect. Computers in particular "can affect several aspects of the test method" (Chapelle, 2001, p.96). In evaluating the speaking skill, test method may be a crucial dimension effecting learner performance. Anxiety producing methods can act as inhibitors and be detrimental to the desired outcome, obtaining a representative sample of student speech under controlled circumstances. Anxiety created by technology and its uses may be detrimental. As computer technologies are becoming more ubiquitous in education, however, their potential as negative affect creators decreases.

Table 2. Qualities of Test Usefulness

Quality	Definition
Reliability	The consistency of the performance reflected in scores
Construct validity	The appropriateness of the inferences made on the basis of test scores
Authenticity	The correspondence of characteristics of the testing activity to characteristics of relevant non-test contexts where language is used
Interactiveness	The expected extent of involvement of the test takers' knowledge and interest and of their communicative language strategies in accomplishing a test task
Positive impact	The positive consequences that a test can have on society and educational systems and on the individuals within the systems (i.e., learners and teachers)
Practicality	The adequacy of the available resources for the design, development, use and evaluation of the test

Source: Bachman and Palmer, 1996

In envisioning an assessment of speaking ability through computer technologies, looking at the characteristics and attributes of computer software in regards to the demands of the oral task is paramount.

Potential and Limitations of Computer Technologies for Oral Assessment

It is impossible at this time to determine the potential and limitations of computer technologies with any degree of precision or certitude. The technology evolves constantly, and the pedagogical applications to these new technologies are ever being developed according to old and new theories of language acquisition and language teaching. The issue, therefore, requires to be looked at from an outcomes point of view, and from there to assess what is needed from the technologies (Zhao, 2003).

As mentioned above, the desired outcome of this study and others of similar breadth is to obtain a representative sample of student speech following a specified curriculum in a practical and authentic manner. In other words, foreign language instructors want to know whether students are adequately progressing in their learning of the communicative features of the language; they need the assessment to be powerful enough and flexible enough that it gives them the opportunity to include multimedia so as to approximate or replicate authentic contexts for communication (quality of authenticity in terms of test usefulness, Bachman & Palmer, Table 2). The tasks mediated through the computer software must be engaging so as to offer learners' opportunities to demonstrate their knowledge and interest in the TL (quality of interactivity in terms of test usefulness, Bachman & Palmer, Table 2); there needs be a permanent recording of the learner's speech generated, so that it may be stored for comparison as well as for delayed evaluation (quality of practicality in terms of test usefulness, Bachman & Palmer, Table 2). The computer interface must be simple to use, both from the instructor's as from the students' end; it must also be easily integrated into a course management system (CMS) (quality of practicality in terms of test usefulness, Bachman & Palmer, Table 2). The

41

possibility of giving the learner oral or written feedback on their performance is also highly recommended. The last essential criterion for a practical test certainly is its cost: it must be free or relatively cheap for an educational environment constantly pressed for more effectiveness with diminishing means.

Computer hardware has greatly evolved in the last decade, and software developers, prodded by classroom instructors and researchers, have made some incursions into designing oral proficiency tests, as was mentioned earlier (COPI, TOEFL, STAMP, etc.). Tests of oral prochievement, however, seem to have attracted less interest, possibly because its advocates have not found their voice. A thorough search for existing software products responding to the above-mentioned criteria has returned few hits. This is on par with the few references in literature since 2000 as to the role of computer technology in the development of oral skill. The reasons advanced for such dearth of references is not lack of interest in developing the oral skill, but merely the fact that the technology necessary to practice and develop the oral skill represent a significant challenge for both hardware and software developers (Barr et al, 2005).

Three software products, freely or cheaply available to the practitioner, have been tested for suitability according to Bachman and Palmer's qualities of test usefulness. The first, developed by Brigham Young University's Dr. Larson, is the OTS (Oral Testing Software). The product is sold and has instructional licenses available for under $100. It is a bi-platform (Mac and Windows), CD-ROM delivered test, which allows instructors to create a test by using some multimedia (pictures, oral prompt) as well as written prompt; the student needs to complete the test on a computer that has been licensed and unto which the CD-ROM has been uploaded. The interface is not intuitive, and has a

steep learning curve for the instructor. It is rather simple for the student to use, but as it is platform-dependent, students must come on campus to use it. This platform-dependence, the licensing, and the lack of intuitiveness made this software unsuitable for the purpose of this study: its possible use for the research project was quickly rejected.

The second software, the LARCstar, was developed by the Language Acquisition Research Center at San Diego University. It is web-delivered, and is totally free for instructors interested in using it (open source freeware). It requires being hosted in a server dedicated to its use and is uniquely available for Windows platforms and servers. Additionally, it requires the enabling of an applet, Active X, that is commonly disabled by computer lab administrators and anti-virus software for its potential in carrying viruses: this was the main drawback of the software as experienced in the pilot that was conducted. The administrative load (entering students' names one by one and creating individual passwords) was heavy, and the multimedia accepted only QuickTime movies, which proved to be another drawback of the product. As this software did not comply either with the quality of practicality advocated by Bachman and Palmer, it was rejected.

The last oral prochievement assessment software found is called RIA (Rich Internet Applications) and has been developed by CLEAR, the Center for Language Education and Research at Michigan State University. The instructor module is web-delivered and web-accessible on the CLEAR server: it is free and only necessitates signing in. Pictures, movies, audio prompts can be uploaded to create interactive tasks, that can subsequently be imported into the instructor's course management system or on a private website; students' voice recordings are stored on the CLEAR server and accessible to the instructor through his/her account. The software is not platform-

dependent, does not require any downloads, is simple to use, and is free for users. As it passed the criterion for practicality and interactivity in terms of Bachman and Palmer's test usefulness, it was deemed suitable for the purpose of the current study project.

Naturally, as Zhao (2003) pointed out, the interface is as good as the task and its uses, and the software itself does not represent the sole criterion for quality in an oral prochievement assessment. Its proffered practicality and its potential for interactivity and authenticity, however, are an essential ingredient to respond to the study's research questions: (1) How does technology affects beginner-level FL instructors' and students' experiences with oral language assessment? And (2) How does a computer-mediated oral assessment facilitate the alignment between the challenges of oral assessment and best practices in assessment? It is the contention of this research that CLEAR's RIAs, although it does not provide all the preferred characteristics for optimum quality of test usefulness, is promising, and may help in answering the questions. Participants to the study will have the opportunity to determine which criteria of Bachman and Palmer's test usefulness they deem essential for an effective computer-mediated oral prochievement test.

Chapter 3

Research Methods

Research and research methodologies are essentially dependent on the perspectives and epistemological traditions that characterize the purpose of the study, the context in which the study is conducted as well as the individual and social perspectives the researcher brings about in her interactions with and understandings of the world. It is the goal of this chapter to offer the reader a rationale of the choice of grounded theory methods to explore and construct an understanding of the ways a technology-mediated oral test can bring congruence between best practices in assessment and the very challenges that assessing the oral skill purport.

Through this chapter the reader will also obtain an appreciation of the selection criteria for the study's participants and software; the perspectives, position, and role adopted by the researcher that will bear on the trustworthiness of the study; as well as the data collection and analysis procedures. This chapter also aims at explicating the contextual correspondence between methodological choices and the following research questions:

Research question 1: How does technology affect beginner-level foreign language instructors' and students' experiences with oral language assessment?

Research question 2: How does a technology-mediated oral assessment impact the alignment between the challenges of oral assessment and best practices in assessment?

Sub-questions:

1. How do instructors of French and Spanish conceptualize the assessment of speaking achievement?

2. How do instructors of French and Spanish experience a technology-mediated assessment of speaking achievement?

3. How do beginner students of French and Spanish experience a technology-mediated assessment of their speaking development?

Design of the Study

The nature of the phenomenon of interest in this study as well as my own dispositions as a researcher have led me to take a constructivist approach to grounded theory such as Charmaz (2003, 2006) conceptualizes. Grounded theory is interested in forming theory substantiated from and grounded in data (Schram, 2006). It consists of "systematic, yet flexible guidelines for collecting and analyzing qualitative data to construct theories 'grounded' in the data themselves" (Charmaz, 2006, p. 3). It particularly reflects the processes or changes that occur over time in actions and interactions among people and events around a substantive topic (Strauss and Corbin, 1998). As Schram (2006) describes it, constructivist grounded theory brings in a

subjective dimension to grounded theory, as it emphasizes the feelings, perceptions, and meaning-making of all study participants, including those of the researcher. Grounded theory does not prescribe particular data collection techniques. The rigor of its analytic process is what makes it powerful, as it aims to develop, refine, and interrelate the concepts under study (Charmaz, 2003). As early data are collected, they are taken apart, classified, identified and synthesized through coding. As patterns of interactions and relationships are analyzed and confronted with new data, a conceptually dense theory is developed, and constantly revised as new data inform the study. Constructivist grounded theory (CGT) proceeds from a few basic assumptions as described in Schram (2003, p. 74):

- "Human beings are purposive agents who take an active role in interpreting and responding to problematic situations rather than simply reacting to experiences and stimuli.

- Persons act on the basis of meaning, and this meaning is defined and redefined through interaction.

- Reality is negotiated between people (that is, socially constructed) and is constantly changing and evolving.

- Central to understanding the evolving nature of events is an awareness of the interrelationships among causes, conditions, and consequences.

- A theory is not the formulation of some discovered aspect of a reality that already exists 'out there'. Rather, theories are provisional and fallible interpretations, limited in time (historically embedded) and constantly in need of qualification.

- Generating theory and doing social research are part of the same process."

47

Selection of Site, Participants, and Software

Participants and Site

The study will take place at a university whose language department currently practices midterm and final oral interviews, or mid- and final equivalent oral assessent, to determine first and second semester students' achievement in oral proficiency. Anecdotal sources as well as my own observations have brought to attention the dissatisfaction that foreign language teachers, at all levels of instruction, feel in the logistics and results of the oral interview as it is practiced. Teachers have expressed on numerous occasions their desire to see the assessment of the speaking skill done in a manner more aligned with current theories of assessment as well as best practices. As the interest to try something different exists in these particular environments, a site such as the language department of a university constitutes excellent ground to conduct educational research.

Five instructors will be the main participants to this study. Participants will be selected on the basis of their interest in assessing the speaking skill and their willingness to try new concepts and experiment with technology. It is anticipated that the teacher participants will be of different educational and experiential backgrounds, and will have various proficiency in using technology in instruction. The foreign languages taught in the various participants' courses, are inconsequential to the study, to the extent that the target language itself is not the object of the study.

I elected to select five participants for this study to ensure that a sufficient pool of three instructors will provide the rich data I seek, considering the real possibilities of attrition over a semester of data collection.

Selected students of French and Spanish will be recruited at the end of the semester to share their experiences with the computer-mediated oral achievement test. One to two students per instructor will be invited to talk about their experiences according to the following a-priori criteria:

Table 3 Criteria for Selection of Student Participants

	Low Use of Software	High Use of Software
Low Performance in Class	☐	☐
Good Performance in Class	☐	☐

It is possible, however, that throughout the semester of data collection different student issues with the computer-mediated oral test arise. If such were the case, and in accordance with constructivist grounded theory, I may elect to modify the above-mentioned criteria for selection of student participation to include students' experiences emerging from use of the software.

Pilot of Software

An extensive search into available software programs was conducted, giving careful consideration to the advantages and disadvantages of different options such as the purpose of the software (testing proficiency vs. achievement), user cost, mode of delivery (web-delivery versus CD-ROM), security, ease of access and use, multimedia and feedback capacity, and data storage location and capacity. This search returned few

results, as technological advances in computer capabilities have been slow in addressing the multidimensionality challenge of the speaking skill.

The LARCstar, a software program developed by San Diego State University's Language Acquisition Resource Center (LARC), seemed most promising, as a web-delivered oral testing program for the purpose of practice and assessment of achievement in the foreign language classroom. LARCstar has multimedia capacity; students' recorded submissions are stored on a server, which allows for long-term storage and differed evaluation; instructors can post written feedback to student submissions. I was able to obtain permission to use it from LARC, and decided to pilot it fall 2007.

When I approached the French Department Coordinator and presented her with the concept of offering students a computer-mediated oral assessment, she was enthusiastic and granted me permission. I piloted the software with one interested teaching assistant and her French 1 class; the results were inconclusive, as the instructor had chosen to present her students with the opportunity to view the computer-mediated oral testing as extra credit: few students took advantage of that opportunity. Seeing that as extra credit the pilot has been partially successful, the French Department Coordinator decided that all classes of lower levels of French would be given the opportunity to benefit from the computer-mediated speaking test the following semester. Thus, beginning Spring 2008, four French 1 instructors and their students (4 classes), and two French 2 instructors and their students (4 classes), or the entire population of students and instructors for the beginning levels of French, started being trained in using the computer-mediated oral assessment. This was my second pilot of the LARCstar software.

The French TA instructors were briefed during the first teaching assistant meeting of the semester (first Thursday after the beginning of classes) that a speaking test was being included in their course requirements (the speaking test had not been expressly included in the French I and II syllabi). A sample computer-assisted speaking test was demonstrated to the instructors during that initial meeting. Subsequent visits in the instructors' different classes were performed during the second and third weeks of classes, explaining and demonstrating to students the functioning of the software program. An explicative brochure was given to the students, detailing the procedures to follow to take the speaking test. Instructors were given instructions as to how to access their students' recordings and how to grade them.

I met with the French TAs in their level groups throughout the semester during their scheduled 'end of unit' meeting time; They would agree on the topic and questions for the upcoming oral prochievement test, let me know what kinds of pictures they wanted for visual support, and I would subsequently construct the test, submit it to them, and upload it with their agreement. The instructors had opted out of the possibility of constructing themselves the oral prochievement test, arguing lack of time and need for consensus. I assumed therefore the role of researcher, technical support, as well as test constructor during this pilot.

Results of the pilot

Because the administration of the software is done through San Diego State University, permission to use the program and to enroll instructors and students is

dependent on the LARCstar administration, and any issue with the program takes some time to be resolved as a consequence. During the two pilots of the software conducted in fall 2007 and spring 2008, I found LARCstar administrators quite responsive and interested in providing a great service to their users. The lag time between a technical issue and its resolution, however, negatively impacted the teachers' and students' perception of effectiveness of the software program.

The LARCstar software has been developed to function exclusively with Windows 2000 or XT, and requires an applet only available on newer versions of Internet Explorer. As a result, non-Windows users, older computer users, and non Internet Explorer users could not use the program. A special permission was obtained from the IT department to make available seven Windows computers in a lab in which the required applet, ordinarily disabled by most virus scanning software, was enabled. These platform and browser restrictions were the biggest hardware/practical limitations of the software program.

Class set up was a lengthy and tedious process, as each student in a class needed to be entered manually. I took the responsibility of creating all the classes for the instructors, and saw first hand that it was a very time-consuming process.

Student IDs and passwords were manually generated when the class was created, as part of their email address, with the possibility to be modified later on by the student; it was the responsibility of the instructor (or me in this instance, as I assume that responsibility) to communicate login ID and password to students, along with the procedure to change their password. As the opportunity for error existed both in the

creation of student login and in the changing of their passwords, many students subsequently experienced difficulty accessing the test because of inaccurate or forgotten login ID and password.

Test creation requires a number of steps that are not intuitive. The visual and audio prompts must be created in a movie making software, uploaded onto the LARCstar teacher module multimedia bank, then uploaded into the test creation module. Some instructors found the process time-consuming and confusing.

Instructors' attitudes toward the project: new instructors were not receptive to what they perceived was a further imposition on their learning to teach load. They felt the learning curve of the software was too steep (complicated, confusing, logistically challenging), and few required of their students that they practice or test for each unit. More experienced instructors were satisfied with the workings of the program, although the hardware/practical limitations encountered prompted several of them to treat the oral assessment, again, as extra credit.

Implications of the Pilot for the Study

The issues of platform-dependence and further restrictions to Internet browser/applet led me to conclude that such software was not ideal for conducting my study. Web-delivery is important, as it allows students to take the oral assessment anywhere anytime, compared to the restricted access that a CD-dependent software would bring to the project. It is, however, preferable to use a software program that is

CD-bound over a web-delivered program when this latter brings more restrictions to access and availability.

Administrative issues such as class and test set up, login and password protection, etc. must be under instructor control as much as possible, and must be simplified to the maximum, as a consideration for instructors and students who are not technology-savvy, yet who could work with simple interfaces.

The test creation process must be simple, logical, and self-explanatory. Instructors are looking for an easy to use, effective and efficient tool that will be as simple and to the point in setting up as it is in evaluating. For student use, simplicity and intuitiveness are also a requirement.

Teacher buy-in is essential. Most teachers of languages agree that assessing speaking development is a time-consuming process and that they would be interested in having at their disposition an assessment that would allow them to obtain rich language samples while not taking class time. The pilot, however, demonstrated that such an instrument, when imposed on unsuspecting instructors, as dedicated and willing as they were, does not create the motivation in instructors necessary to, in turn, motivate students. Moreover, when technology is perceived as adding stress and workload, when it is not functioning up to expectations, it becomes detrimental to the project goal, which is to encourage instructors and students to practice and assess the oral skill.

I contemplated using the Brigham Young University OTS, a CD-delivered program of oral ability assessment. It is available for both Windows and Macintosh platforms, yet it means that the program needs to be installed in a certain number of

computers in a lab for example, making it less accessible to students and restricting their availability. It is a little confusing to set up, as the interface is busy and not intuitive. For reasons similar to those expressed above, I decided against using this particular software for my study, although it is a good product and may be interesting to technology-savvy users.

Rich Internet Applications for Language Learning (RIA)

The Rich Internet Applications developed by the Center for Language Education and Research (CLEAR) of Michigan State University combine several of the features that are recommended for my study. They are web-delivered and are not platform dependent; they are stored on the CLEAR server with an instructor login, yet are made available to the students on any website chosen by the instructor; they are free, simple and intuitive to use. They offer multimedia capacity; they can be as simple as a recording audio box, in which students record the language that they practice; they can also be a simulated interview, in which teacher records questions (with possibility of including a visual prompt), students asynchronously access the questions, listen, and record their answers, and teacher accesses students' recordings later on for evaluation. Instructors have total control of all administrative and delivery options, which are simple to create and implement. I will therefore suggest to my participant teachers to use the RIA applications for the study, yet will remain open to their suggestions if they prefer using another software program, as I am not evaluating a particular software program, but am aware that satisfaction with and motivation to assess oral development depends in large part in the practicality and usability of the technology tool used.

Figure 2 Example of a Student Practice Activity with Audio Dropbox on a class website

Instructors will be given the opportunity, through the CLEAR Audio Dropboxes,

to have novice low to novice mid students practice daily or weekly specific features

under study. Figure 3.1 illustrates one type of tasks that can become routine in the foreign language classroom: it merely requires students to introduce themselves, by giving their first and last names, and their age. These features would have been seen and practiced in class, and the audio dropbox gives students further opportunity to practice the specific features. Instructions are given as to how to practice and record one's speech; the simple rubric with which the speech sample will be evaluated, places the emphasis on making speech comprehensible, in accordance with Swain's theory of forced output (1985). By having students practice these features daily, bi-weekly or weekly, and giving them the opportunity to record themselves until they are satisfied with their speech, the speaking skill takes on a greater importance, and students are more motivated. Instructors can easily access students' recordings (see Figure 3.2) and quickly assess them according to the simple rubric. There is not yet a feedback function, but this function, according to my conversation with the RIA developers, is in the works and should be available in the fall of 2009 (personal communication with Dr. Dennie Hoopingarner, April 1, 2009).

Conversations RIA appeals to both listening and speaking skills, in that it requires students to listen to teachers' voice-recorded question(s), before recording their answers. It simulates a conversation, or an interview such as is routinely practiced during midterms and finals at the lower levels of proficiency, where actual interaction is reduced. Students have here the opportunity to listen to the instructor's questions in practice mode, where they can listen to the audio prompts many times, and record themselves until satisfied with their speech. Or they can (an instructor's option) listen to the questions in real-time, where recording starts automatically after the audio prompt has been uttered. This latter option could be kept for the midterm or final speaking assessment, if students have been

practicing all semester long at the end of each unit of instruction for example, with 'conversations tests'. Practice and assessment are here aligned, as students are familiar with the task as well as the task format.

Figure 3 Example of Student Submissions in Teacher Audio Dropboxes Site

Figure 4 Examples of Student Submissions in Teacher Conversations Site
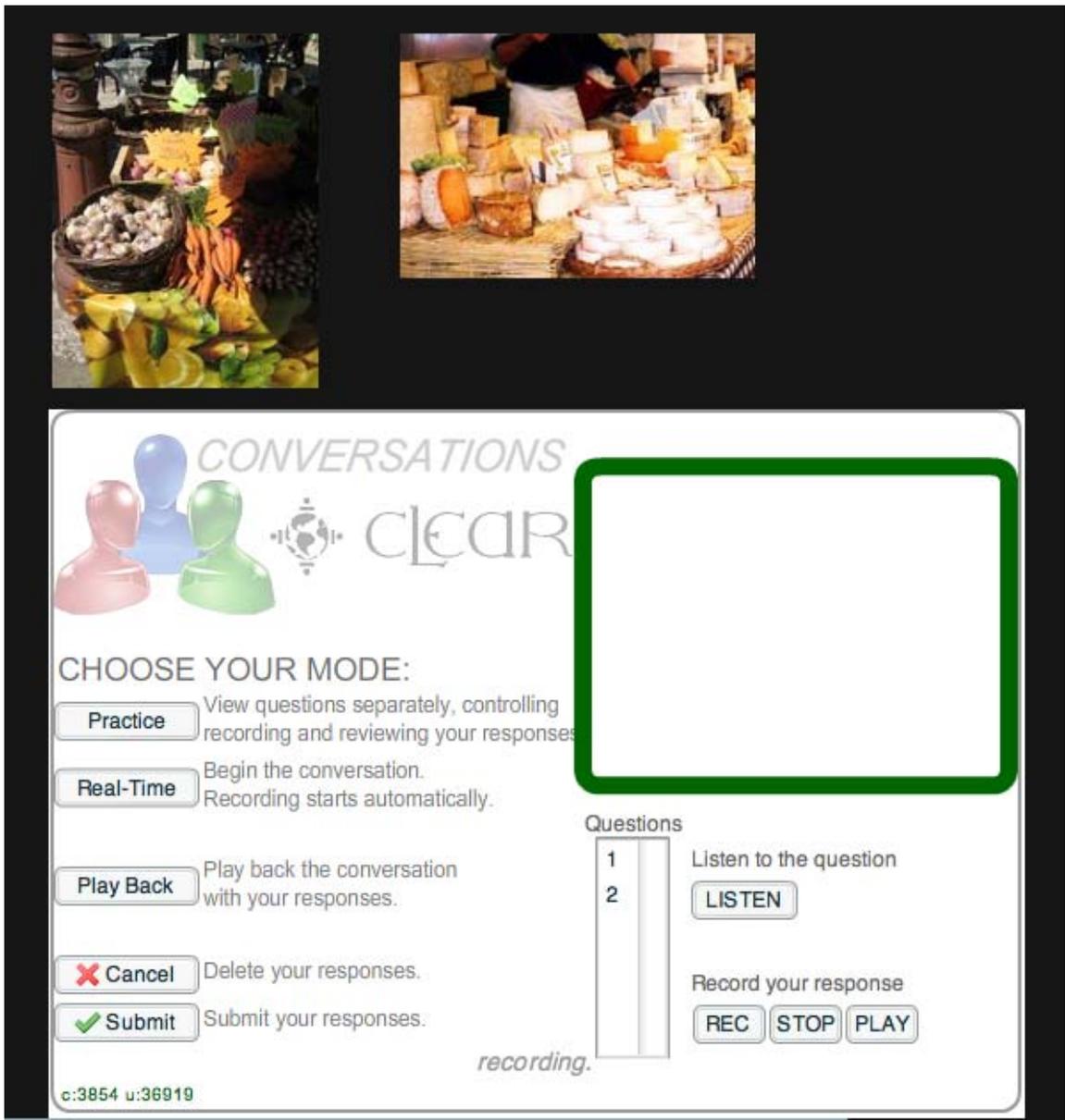
Figure 5 Example of test with practice activity on a class website

I will provide training for instructors and students in the use of the technological tool, and will also provide support for any need that may emerge during the study. CLEAR RIAs are available at http://clear.msu.edu/teaching/online/ria/ and are free for all to use.

Background, Stance and Trustworthiness

In order to ensure the credibility of the research endeavor and its trustworthiness to its participants as well as its readers, it is essential for the researcher in a qualitative study to set out having defined her role and understood her biases. Acknowledgment of the subjectivity inherent to doing research in social sciences lends credence to assertions made from choices and interpretations based on data. A focused inquiry means that stances have been taken, preconceptions formed. It is important to keep this subjectivity, these preconceptions in mind and question them as data collection and analyses are performed. By engaging and monitoring one's subjectivity, one takes feelings and emotions prompted by events or participants' responses into account to probe at the reasons for interpretations, as well as the implications of such interpretations of data (Glesne, 1999; LeCompte, Schensul, Weeks, & Singer, 1999; Peshkin, 1988). I will therefore make explicit here my background and stance in order to expose my preconceptions as I enter this inquiry.

I have taught French 1 and 2 in the department of World Language Education and was dissatisfied at the time with the way oral assessment was conducted. I had the same experience in the foreign language classes I taught at the middle and high school levels (Spanish and French). I found the practice of once or twice a semester oral interview to be inefficient and a difficult endeavor for both teachers and students. Many of my fellow instructors at the university and secondary levels had the same issue with logistics and

time, to the extent that some would renounce the practice of oral assessment altogether. These experiences greatly influenced the choice of this research topic. I am interested in practicing teaching and assessment in a seamless continuum, as teaching and assessing are two aspects of the same instructional act. I am passionate about bringing oral achievement in the classroom to the same level of accessibility and reliability that other language skills know, and I strongly believe that technology can be a significant tool toward making the assessment of oral development in the foreign language classroom a more practical, efficient, and satisfactory endeavor for students and instructors alike.

I am aware, however, that all instructors and students may not share my points of view about the assessment of speaking ability and achievement. My position comes from years of experience teaching, as well as my studies in second language acquisition, learning and teaching. Instructors with different backgrounds, experiences, and interests might not see the assessment of speaking as essential as I do, or the ways and means to obtain valid and reliable information as important as I do. I will therefore inquire about teachers' understandings of best practices in assessment in order to adequately inform the study. Moreover, I am very conscious of the fact that interest in instructional technology and savvyness in technology use may play an important role in instructors' and students' appreciation and understanding of a computer-mediated speaking assessment. These issues are some of the preconceptions I bring to the study, and will be dealt with through careful and systematic analysis of data, as well as through member checking. I will also engage another researcher used to conducting qualitative inquiry in analyzing my data, to ensure that the codes, categories and theories remain grounded in data. Original

transcripts, coding schemes and memos will be made available as well to ensure credibility.

The ethicality of my research, as it concerns the integrity of my relationship with my study participants, will be ensured through careful consideration and safeguards as put forth by the Institutional Review Board of this institution.

Researcher's Role

I will assume the role of participant observer in this study. This role, according to Patton (1990), requires direct participation and observation along with interviewing. Through this role I will obtain an insider's view of the events occurring during instructors' interactions with the software. I will interact with the instructors throughout the semester as technology support for the oral assessment. I will work closely with them to determine and create the content of each oral assessment; as such I will directly participate in the process of conceptualizing the oral tests, with a direct handle on the issues the instructors experience with the oral assessment. I will also act as observant and resource to them in their grading efforts, as their perceptions of the usefulness of the computer-mediated oral prochievement test may be expressed at that time. I will also be providing support to both instructors and students via email if required.

Data Collection Procedures

Grounded theory methods do not prescribe particular data collection techniques. Observations, interactions, materials gathering, interviews, are all possible data collection techniques. Most importantly when collecting data, according to Charmaz (2006, p.3), "[grounded theorists] begin by being open to what is happening in the studied scenes and interview statements so that we might learn about our research participants' lives. […] Grounded theorists start with data. We construct these data through our observations, interactions, and materials that we gather about the topic or setting. We study empirical events and experiences and pursue our hunches and potential analytic ideas about them."

Data for this study will be collected throughout a 15-week semester during individual interviews with and observation of the selected instructors, my main participants, as well as, at the end of the semester, interviews and/or discussion panels with selected students.

Individual Interviews with Instructors

One initial interview will be conducted with all instructors consenting to participate in the study, and other interviews will be scheduled over the course of the semester and the following months as deemed necessary by the emergence of theoretical categories from the data analysis, and the need to confirm these categories (theoretical sampling). Interviews will be audio-recorded, transcribed, and analyzed.

Observations of Instructors' interactions with the software

I will meet with each instructor at the beginning of the semester to present the software, and will offer them my technical assistance to help create the assessments and evaluate them over the semester. I will not, however, create the assessments for them, as, as the pilot demonstrated, it did not create ownership of the assessments in instructors, which in turn, did not create interest in the student. I anticipate, depending on instructor confidence in using and familiarity with teaching with technology, to provide more help to some instructors than others. During these help sessions, I will observe the instructors' interaction with the software. Verbal and non-verbal interactions will be audio-recorded and/or logged in a field journal, transcribed and subsequently analyzed. Additionally, any email interactions will be logged in and added to the data being collected.

Interviews and/or Panel Discussions with Selected Students

Toward the end of the semester, after the final interview, 1 to 2 students per instructor will be selected to participate in individual interviews and/or panel discussions, according to the criteria above-mentioned, or emerging issues in their interaction with the software. During these sessions, students' perceptions of their oral development through computer technology will be elicited. The discussions and interviews will be audio-recorded, transcribed, and subsequently analyzed.

Additionally, I will keep a log of my observations and my thoughts while collecting data during all three types of above-mentioned events.

Data Analysis Procedures

The data collected will be analyzed following the systematic procedures inherent to the constructivist grounded theory. The practice of grounded theory comprises:

- "Simultaneous collection and analysis of data

- A two-step data coding procedure

- Comparative methods

- Memo writing aimed at the construction of conceptual analyses

- Sampling to refine the researcher's emerging theoretical ideas,

- Integration of the theoretical framework"

(Charmaz, 2003, p. 251)

Grounded theory then, as defined by Glaser and Strauss (1967) and Glaser (1978, 1992), involves concurrent data collection and analysis. To do so, constant comparative methods are used, according to which new data are continually being compared to existing data, and this at each level of analysis. This systematic procedure is cyclical, in that new data is constantly analyzed against the categories that emerged from analysis and synthesis of earlier data. The process of analysis is illustrated in Figure 3.1.

The very first data collected, which I anticipate to be the initial interviews with participant instructors, will be immediately submitted to line-by-line coding. By studying these data closely, literally line by line, concepts relative to instructors' experiences with and expectations for the computer-assisted speaking test will emerge: this will allow me to conceptualize the ideas generated by the data and express them into labels. From these labels, I will write memos, which are extensive notes on the codes and their immediate correspondence with the data. At this point more data will be collected, which in my study could be an incidental conversation with a student or an instructor, a TA meeting in which the speaking test is discussed, or an assessment session with an instructor.

An essential quality of constructivist grounded theory is that data should not be forced into preconceived categories. "Theoretical categories must be developed from analysis of the collected data and must fit them; these categories must explain the data they subsume. […] Any existing concept must earn its way into the analysis" (Charmaz, 2003, p.251). Although my personal experiences of the process of testing the speaking achievement of beginner students of French and Spanish and my review of the literature have necessarily led me to formulate ideas and categories for data analysis and anticipate possible issues to be raised during my data collection, it is the nature of grounded theory practice to let categories emerge from the data and to seek confirmation of these new categories through theoretical sampling. Charmaz differentiates theoretical sampling from initial sampling. Initial sampling consists initial data collection meant to answer research questions as well as access: it is a start for data collection.

Writing the first draft

Further theoretical sampling if needed

Integrating memos diagramming concepts

Theoretical memo-writing & further refining of concepts

Sorting memos

Adopting certain categories as theoretical concepts

Theoretical sampling Seek specific new data

Advanced memos refining conceptual categories

Data collection --------- Focused coding

Initial memos, raising codes to tentative categories

Initial coding   data collection

Sensitizing concepts & general disciplinary perspectives

Reexamination of earlier data

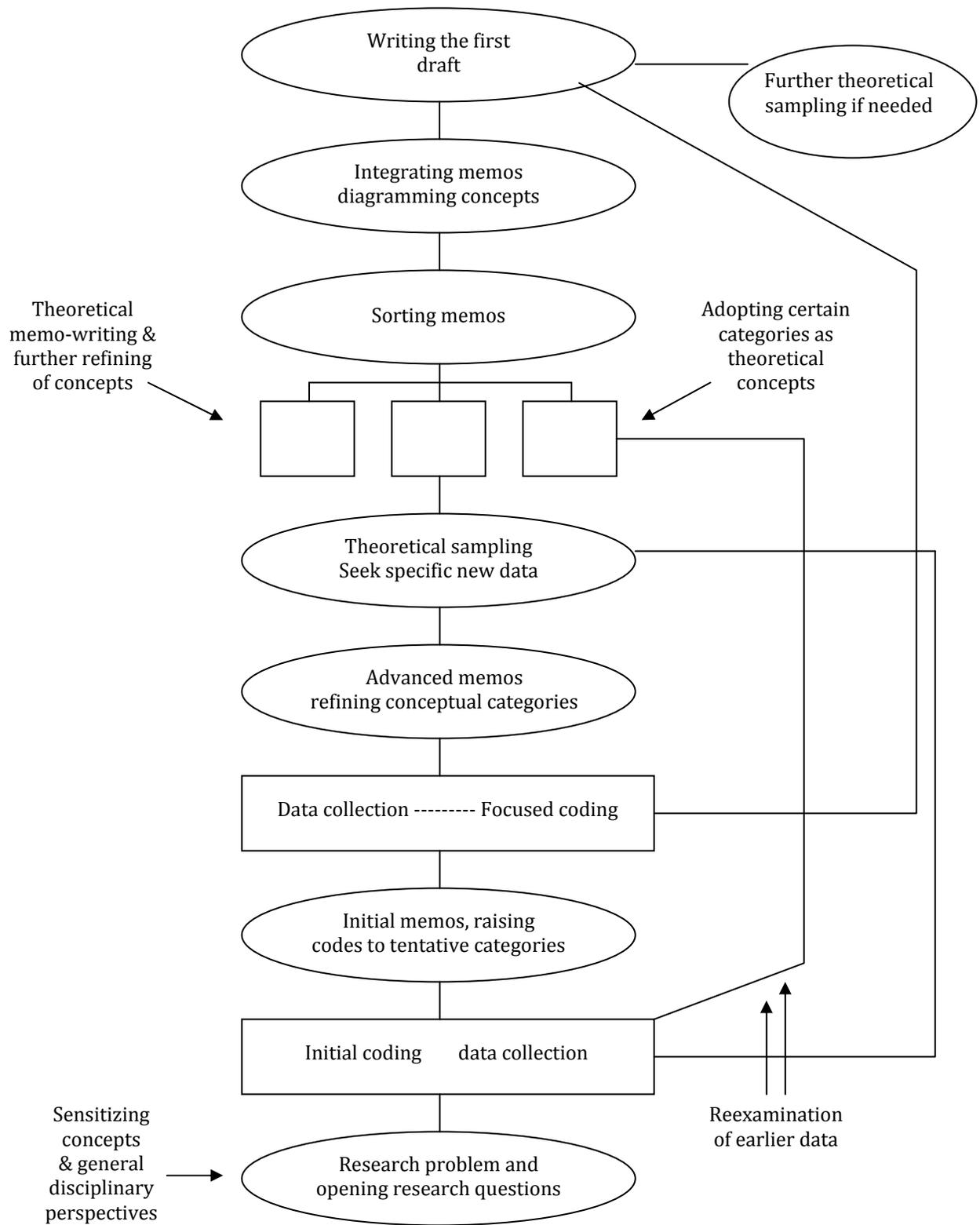Research problem and opening research questions

Figure 6 The grounded theory process (Charmaz, 2006, p.11)

Once the researcher has started analyzing data and constructed categories, s/he moves on to theoretical sampling, which takes on a rather different logic and meaning than theoretical sampling as understood in quantitative inquiry or even other qualitative endeavors. In grounded theory, theoretical sampling seeks to obtain further data to help explicate categories. Participants, interactions and events are questioned anew to see if more data can be collected to fill these categories. If categories remain thin and further data cannot fill them, these categories and line of reasoning are abandoned. On the other end, "When your categories are full, they reflect qualities of your respondents' experiences and provide a useful analytic handle for understanding them. In short, theoretical sampling pertains only to conceptual and theoretical development; it is *not* about representing a population or increasing the statistical generalizability of [one's] results." (Charmaz, 2006, p.100, italics in the original text). Category saturation occurs when new data no longer bring new theoretical understandings or distinguishing features to these core theoretical categories.

The Dynamic Cycle of Data Analysis Procedures

1. Line-by-line coding: going from concrete statement to analytic interpretation through categorization. Each line or segment of data will be given a short name or label that will summarize and account for each piece of the data under analysis.

2. Focused coding: selecting, sorting, synthesizing, integrating and organizing the codes issued from the line-by-line coding into the most salient categories representing the data.

3. Theoretical sampling: gathering further data to fill gap in categories, and refine categories boundaries. This sampling jumpstarts the cycle of line-by-line coding and focused coding (analytic phase). Charmaz (2003) advocates performing theoretical sampling late into the analysis to enable natural emergence of analytic directions, as well as to prevent premature closing of the analysis.

4. Theoretical coding: specifying the possible relationships between the categories developed during focused coding.

5. Memo-writing: making written comparisons between and among data, codes, categories, concepts, and articulating inferences about these comparisons.

6. Theoretical sampling: gathering further data to fill gap in categories, and reach data saturation, starting another analytic phase.

7. Theoretical sorting, diagramming, and integrating: creating and refining theoretical links and relationships to form emerging theory.

References

Abd-Kadir, J. & Hardman, F. (2007). The discourse of whole class teaching: a comparative study of Kenyan and Nigerian primary English lessons. *Language and Education 21*(1), 1–15.

American Council on the Teaching of Foreign Languages (1999). *ACTFL Proficiency guidelines—speaking: Revised 1999.* Retrieved May 3, 2005, from http://www.actfl.org/files/public/guidelinesspeak.pdf

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Barnwell, D.P. (1996). A history of foreign language testing in the United States: From its beginnings to the present. Tempe, AZ: Bilingual Press.

Barr, D., Leaky. J., & Ranchoux, A. (2005). Told like it is! An evaluation of an integrated oral development pilot project. *Language Learning and Technology, 9* (3), 55-78.

Blaz, D. (2000). A collection of performance tasks and rubrics. Larchmont, NY: Eye on Education.

Breiner-Sanders, K.E., Lowe, Jr., P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines—speaking, revised. *Foreign Language Annals, 33,*13-18.

Brown, H.D. (1973). Affective variables in second language acquisition. *Language Learning, 23,*231-244.

Brown, H.D. (1994). *Principles of language learning and teaching.* 3rd edition. Englewood Cliffs, NJ: Prentice Hall Regents.

Burns, A. & Joyce, H. (1997). *Focus on speaking.* Sydney: National Centre for English Language Teaching and Research.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1-47.

Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N.K. Denzin & Y.S. Lincoln, (eds.), *Strategies of qualitative Inquiry,* (pp. 249-291). Thousand Oaks, CA: Sage.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.

Chamot, A. (1987). The learning strategies of ESL students. In A. Wenden & J. Rubin (eds.), *Learner strategies in language learning.* New York: Prentice Hall.

Chapelle, C.A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace, Jovanovich.

Corder, S.P. (1973). *Introducing applied linguistics.* Harmondworth, UK: Penguin.

Crookes, G. (1991). Second language speech production research: A methodological

    oriented review. *Studies in Second Language Acquisition, 13*(2), 113-32.

Dickinson, L. (1987). *Self-instruction in language learning.* Cambridge: Cambridge

    University Press.

Egan, K. (1999). Speaking: A critical skill and challenge. *CALICO Journal, 16*, 277-293.

Ellis, (1992). Second language acquisition & language pedagogy. Philadelphia :

    Multilingual Matters.

Firth, A. & Wagner, J. (1997). On discourse, communication, and (some) fundamental

    concepts in SLA research. *The Modern Language Journal, 82,*91-94.

Firth, A. & Wagner, J. (2007). Second/foreign language learning as a social

    accomplishment: Elaborations on a reconceptualized SLA. *The Modern Language*

    *Journal, 91,*Focus Issue, 800-819.

Flewelling, J. (2002). From language lab to multimedia lab: oral language assessment in

    the new millennium. In C.M. Cherry (Ed.) *Dimension 2002: Cyberspace and*

    Foreign Languages: Making the Connection. Valdosta, GA. 33-42

Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown & S. O.

    Yamashita (eds.), *Language testing in Japan* (pp. 101-106). Tokyo: Japan

    Association for Language Teaching.

Glaser, B.G. (1978). *Theoretical sensitivity*. Mill Valley: Sociology Press.

Glaser, B.G. (1992). *Basics of grounded theory analysis: Emergence vs. forcing.* Mill

      Valley: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for*

      qualitative research. New York: Aldine Publishing Company.

Glesne, C. (1999). Becoming qualitative researchers: an introduction. New York:

      Longman.

Gonzales Pino, B. (1989). Prochievement testing of speaking. *Foreign Language Annals,*

      22, 487-96.

Hall, J.K. (1997). A consideration of SLA as a theory of practice: A response to Firth and

      Wagner. *The Modern Language Journal, 81,* 301-306.

Holec, H. (1985). On autonomy: Some elementary concepts. In P. Riley (Ed.), *Discourse*

      and learning (pp.173-190). London: Longman.

Hoopingarner, D. (2007). "Listening and Speaking Online." *Proceedings from the 5th*

      International Conference on Internet Chinese Education, pp 159 - 165. Retrieved

      from the Internet June 2, 2009,

      https://www.msu.edu/user/hooping4/ICICE2007.pdf

Horwitz, E.K., Horwitz, M.B. & Cope, J. (1986). Foreign language classroom anxiety.

The Modern Language Journal, 20(2), 123-132.

Hugues, A. (1989). *Testing for language teachers.* Cambridge and New York: Cambridge

      University Press.

Hunt, E. (1987). Science, technology and intelligence. In R.R. Ronning, J.A. Glover, J.C. Conoley, & J.C. Witt (eds.), *The Influence of Cognitive Psychology on Testing* (pp.11-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hymes, (1972). Overcoming blocks to change: A source of spirit and spunk! *Childhood Education, 49*(3), 114-7.

Johnson, (2001). *The art of nonconversation*. New Haven, CT: Yale University Press.

Krashen, S.D. (1982). *Principles and Practice in Second Language Acquisition.* New York: Pergamon.

LeCompte, M.D., Schensul, J.J., Weeks, M.R. & Singer, M. (1999). Researcher Roles & Research Partnerships. In J.J. Schensul & M.D. LeCompte (eds.), *Ethnographer's Toolkit*, Vol. 6. Walnut Creek, CA: Altamira Press.

Liming, Y. (1990). The comprehensible output hypothesis and self-directed learning: A learner's perspective. *TESL Canada Journal, 8*(1), 9-26.

Long, (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*(2), 177-93.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment. *Educational Researcher, 20*(8), 15-21.

Magnan, S. S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals, 19*(5), 429-38.

McCarthy, M., & O'Keeffe, A. (2004). Research in the teaching of speaking. *Annual Review of Applied Linguistics, 24,* 26-43.

Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (eds.), *Test theory for a new generation of tests* (pp. 19-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59,* 439-483.

Nitko, A. (1989) Designing tests that are integrated with instruction. In R.L. Linn (Ed.) *Educational Measurement* (3rd ed). Macmillan, New York. 447-474.

Omaggio Hadley, A. (2001). *Teaching Language in Context* (3rd ed.). Boston, MA: Heinle & Heinle.

Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.). Newbury Park, CA: Sage Publications, Inc.

Payne, J. S. & Ross, B. M. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning and Technology, 9* (3), 33-54.

Perret, G. (1990). The language testing interview: A reappraisal. In J.H. de Jong and D.K. Stevenson (eds.), *Individualizing the assessment of language abilities,* (pp. 225-38). Philadelphia: Multilingual Matters.

Peshkin, A. (1988*).* In search of subjectivity-One's own. *Educational Researcher. 17*(7), 17-21.

Rubin, D. (1987). Divergence and convergence between oral and written communication. *Topics in Language Disorders, 7*(4), 1-18.

Salaberry, R.M. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal, 85*, 39-56.

Savignon, S.J. (1972). *Communicative competence: An experiment in foreign language teaching.* Philadelphia: Center for Curriculum Development.

Schram, T.H. (2003). *Conceptualizing qualitative inquiry: Mindwork for fieldwork in education and the social sciences*. Upper Saddle River, NJ: Merrill Prentice Hall.

Schulz, Renate A. (1986). From achievement to proficiency through classroom instruction: Some caveats." *Modern Language Journal, 70*, 373–79.

Shrum, J.L., & Glisan, E.W. (2010). *Teacher's Handbook: Contextualized Language Instruction* (4th ed.). Boston, MA: Heinle.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory (*2nd ed.). Thousand Oaks, CA: Sage.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass and C. Madden (eds.), *Input in second language acquisition*, (pp. 235-256). New York: Newbury House.

Swain, M. (1995) Three functions of output in second language learning. In G. Cook & B. Seidelhofer (eds.). *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson*, (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.),

    Handbook on research in second language teaching and learning (pp. 471-484).

    Mahwah, NJ: Lawrence Erlbaum.

Swain, M. & Deters, P. (2007). "New" mainstream SLA theory: Expanded and enriched.

    The Modern Language Journal, 91, Focus Issue, 820-836.

Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL*

    Quarterly, 15(3), 285-95.

Tobias, S. (1986). Anxiety and cognitive processes of instruction. In R. Schwarzer (Ed.)

    Self-related cognition in anxiety and motivation. Hilsdale, New Jersey: Lawrence

    Erlbaum Associates, 35-54.

Zhao, Y. (2003). Recent developments in technology and language learning: A literature

    review and meta-analysis. *CALICO Journal, 21*(1), pp.7-27.