STARS

Electronic Theses and Dissertations, 2004-2019

2010

Foreign Language Oral Assessment Practices In Florida Middle And High Schools

Grace Kerr Kellermeier
University of Central Florida

Showcase of Text, Archives, Research & Scholarship

FOREIGN LANGUAGE ORAL ASSESSMENT PRACTICES
IN FLORIDA MIDDLE AND HIGH SCHOOLS

by

GRACE KERR KELLERMEIER
Bachelor of Arts, Rutgers University, 1988
Masters of Education, Stetson University, 2003

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Education
in the College of Education
at the University of Central Florida
Orlando, Florida

Fall Term
2010

Major Professors: Joyce Nutta
                  Stephen Sivo

ABSTRACT

Current foreign language pedagogy encourages a communicative approach to language learning. Instead of the reading and writing focus of the past, this communicative approach requires instruction in all skills, which include reading, writing, listening, speaking and having knowledge of associated cultures. A random sample of Florida foreign language teachers was surveyed to establish whether or not the goals and actual instructional practices were aligned. Respondents were asked to identify how much of a typical class period is dedicated to each of the five skills and how much of a typical unit test is dedicated to the same five skills.

The findings showed that only the instruction and assessment of writing were aligned. Instruction and assessment of the other skills were unequal. A comparison of the means revealed that listening was actually instructed much more than assessed. The other means were similar, including writing. Reading was found to be the only skill that was assessed more than instructed. The variables examined in this study included the level of fluency, level of education, amount of experience of the teacher, available resources, amount of the target language used in the classroom, as well as demographic information. The interaction of the level of education and experience of the teacher was significant. Teacher gender was also significant, although the disparate gender groups made it difficult to compare means. The other variables revealed no statistical significance. Slightly less than half of the respondents stated that they do not believe that they include enough oral assessment as a part of instruction. They reported a lack of time, student resistance, class size, and unreliable technology as reasons for not including what they perceived to be enough oral assessment.

ACKNOWLEDGMENTS

In memory of Mum, who taught me how to be a good student.

To my Dad, who taught me the value of speaking to others in their own languages.

To Joe, for his loving support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Former foreign language students often complain "I took French for two years in high school and can't speak a word." This is perhaps due to language atrophy from non-use of the language after formal instruction. It might be from an insufficient length of instruction to retain the learning or even the result of a foreign language curriculum that is not aligned with the program goals. Foreign language is a component of the earliest curricula in American education, although its purpose and pedagogy have changed significantly.

A History of Foreign Language Curriculum

In the United States of America, the earliest foreign language curriculum appeared in the Latin grammar school. This was the epitome of the academic curriculum, a school for the elite, in terms of scholarliness, wealth, and position. Foreign language was a significant part of the general curriculum since the classic languages of Latin and Greek were required. The teaching method in foreign language instruction was grammar-translation, a paper-and-pencil method of sentence dissection and translation to or from the target language. The focus of this foreign language curriculum was reading and writing. The assessment was linguistic accuracy. Once university studies were conducted using English, there was little practical need for oral or aural fluency. The classics continued to be valued for their mental training and discipline. The study of those languages has continued to be an indicator of an academic curriculum and presumes a higher intellect.

Starting in the 1950s, the audio-lingual method gained popularity in foreign language curriculum. This included an aural prompt by either a tape or a teacher, requiring an oral response by the student. This curriculum was popular at the time of B.F.Skinner and the behaviorists and was similar to that used in U.S. military language training at the time. It involved repetitive drill and practice and could be accomplished by many students, simultaneously. The goal was oral and aural conditioning, rather than the reading and writing focus of the past, and included the living languages of French, German, and Spanish. This method met resistance from cognitive language proponents, such as Noam Chomsky, based on the regurgitation effect of stimulus-response, with questionable comprehension. As late as the 1960s, it was common for reading and writing in the foreign language not to be taught for the first three years of instruction, for fear of native sound-symbol interference (Curtain & Dahlberg, 2004.) It has since been learned that this absence of visual stimulus is counter-productive and that many second language learners find security from having the written language available (Curtain & Dahlberg, 2004.)

In 1973, James Asher introduced an audio-motor approach, which has since gained momentum as Total Physical Response (TPR). The teacher makes commands in the target language, and students follow with the requested action. The target language is "chunked" rather than deciphered, word by word, and allows students to have a silent period similar to that of first language acquisition. Asher, Kusado, and de la Torre (1974) claimed that the application of movement also aids in organizing and assimilating language. In 1993, Asher described this phenomenon as brain-switching, where the right brain detects the incoming data and the left brain is the editor, deciding what should be done about the new information. He claimed that this

movement from comprehension to application is necessary for long-term retention. After a personal silent period, rather than a required immediate oral response, he proposed that students would produce language orally. Transfer to reading and writing skills would follow later in the instructional period. There can be difficulty in transfer to reading and writing in non-phonetic languages, of course, as well as languages which use different characters than the base language. The goal of the audio-motor approach is aural comprehension of a living language which would ultimately promote the other language skills.

The goals of foreign language learners have not changed significantly. Generally, learners seek to communicate when studying, traveling, marrying, or emigrating. Foreign language acquisition often facilitates the socialization or acculturation process, cultural homogenization, or economic globalization (Doughty & Long, 2003).

Combining the Latin grammar school academic focus on reading and writing with the audio-lingual and audio-motor focus on aural comprehension and oral production, the American Council for the Teaching of Foreign Language (ACTFL) introduced its requirements for proficiency in 1982. ACTFL identified provisional proficiency in foreign language as speaking, listening, reading, writing and being familiar with relevant cultures (Gasparro-Liskin, 1985). The ACTFL Oral Proficiency Interview (OPI) has been considered to be the most reliable means by which to establish the existing level of proficiency for placement into the appropriate target language class (Thompson, 1995). Proficiency in speaking, reading, writing, listening, and awareness of cultures has become the goal of the foreign language curriculum. There are many sound pedagogies being applied in foreign language classrooms that facilitate the attainment of these goals.

Student interaction through dialogues has become popular in foreign language curriculum. Students engage in guided conversations, speaking activities, social interaction, and exploratory language. Swain (2001) valued student dialogues for the socio-cultural setting and shared knowledge-building among students. She also saw the opportunity to gain insight into the language acquisition processes. She claimed that, in a socio-cultural setting, learners encounter linguistic problems and attempt to solve them together, collaborating and mediating the language acquisition process (Swain, 2002). Ellis' (2003) task-based language learning also included directed communicative tasks, plus listening for comprehension and learning (receiving direction) and focused versus unfocused tasks, all aiming for a balance of fluency and accuracy in language production.

There exists considerable research on how to best accomplish language production. There has been a debate between the value of comprehensible input (Krashen, 1987) and comprehensible output (Swain & Lapkin, 1995). Krashen's (1987) input hypothesis stated that students must always comprehend what is being said in order to learn a language. Students can be led to understand by context, visual cues, or even a swift translation to base language. Krashen (1987) believed that no learning will occur if the student becomes confused and frustrated, raising his affective filter. Krashen's (1987) monitor hypothesis stated that successful language learners monitor their language production to some degree by applying grammatical structure and rules, correcting their outputs as necessary. Some learners use too much or too little monitor to the detriment of successful communication (Krashen, 1987). Those who have subscribed to the monitor hypothesis are concerned with student affective safety in the foreign language classroom, allowing them to take time and produce when ready so as to ensure

comprehension of language being presented. Krashen (1987), like Asher (1977), encouraged a silent period in which students listen and produce language when ready, simulating natural (first) language acquisition. Asher (1977) claimed that children and adult learners who have 12 initial hours of aural input outperform traditional drill and practice students. Proponents of the monitor hypothesis have not been concerned with initial accuracy. Krashen (1987) visualized language learning as a process of active, trial-and-error construction of language. Swain (1995), on the other hand, claimed that language production and explicit correction and feedback, are critical to proficiency. Krashen (1987) and Nikolov (1998) claimed that students who are taught using communication-based instruction, with less insistence on grammatical accuracy, become more fluent and more accurate in grammatical constructions than do their grammar-based instruction counterparts.

Not concerned with fluency versus accuracy, Wallace (2003) suggested a critical discourse analysis in foreign language curriculum. This called for students to read critically, analyzing and discussing everyday texts, and to use the text as a background or platform from which to engage in language use. Students are alerted to how things are said, not just what is said. Rather than focusing simply on reading, writing, speaking, and listening in a target language, students' overall language abilities are enhanced, resulting in more powerful language use. The goal of this curriculum is for language to be learned, not acquired, and to be examined critically. In this curriculum, the acquisition of language is subconscious and intuitive, whereas the learning of language suggests the application of structure and rules. The classroom is an interpretive community which fosters collaboration and builds relationships.

Immersion is a system of language instruction that requires the teacher to use only target language. Students are expected to use minimal base language, communicating in the target language. In his initial study in 1979, Capretz suggested that video could be used to demonstrate authentic, target language to students, circumventing the boredom and difficulty associated with aural comprehension in target language. Capretz' introduction of the "French in Action" video series in 1989 included scenes that took place in France. These provided authentic aural and cultural exposure. The video was recorded entirely in the target language and provided multiple examples of new terminology, using clips from many international French language films, commercials, and staged conversations. Students were expected to derive meaning from these edited examples, and Capretz did not endorse any manner of translation, claiming that a language cannot ever be truly translated. Swain (2004) was also a proponent of immersion in foreign language, stating that immersion allows teachers and researchers to see how learners use existing knowledge to build new knowledge about the language. According to Swain (2000), 30 years of immersion programs in Canada have demonstrated the importance of learner output, the effectiveness of corrective feedback, the roles of base-target language in group tasks, and age differences in cognitive style. In 2005, however, Swain stated that new research findings suggested that students benefit from the use of their base languages to aid in comprehension. She concluded that the use of multiple L1's (first or base language) inside the immersion classroom was warranted, resulting in a multi-lingual classroom.

Bilingual education has allowed for teachers and students to use both the base and the target languages. The aim has been to use an increasing amount of target language, but comprehension is the key. In a 2001 study, Ozerk and Krashen found that Urdu- or Turkish-

speaking students, who were being taught in Norwegian schools with the use of both languages, learned more than those who were being taught using only Norwegian. Monolingual Norwegian students learned the same amount of the subject matter whether they were learning in a Norwegian-only or bilingual classroom, so foreign students' needs were being met without sacrificing the needs of native speakers.

Although researchers have supported the use of multiple languages in a foreign language classroom, there have been some difficulties. Teachers might not speak more than the base language. If they speaks more than one language, students of other languages might be further isolated. As people migrate to different countries, schools and/or governments attempt to accommodate the linguistic needs of their students. In Florida, any teacher who teaches a non-native-English-speaking student must follow one of several English as a Second or Other Language (ESOL) courses of study depending on the linguistic requirements for comprehension of the subject being taught. Physical education teachers have been required to complete the shortest course, for example, whereas English teachers have been required to complete the longest. Instruction continues in English language immersion with suggested visual modifications. Depending on the needs of the population, some schools have created units to teach English as a Foreign Language (EFL). There has been a significant variety in accommodations made for students with linguistic barriers. Mitchell and Brumfit (1997) found that European students of other than European origin were not being given equal learning opportunities in spite of a series of education reform acts in Europe. Individual teachers, not national curriculum, were determined to accommodate these students.

Increasingly, students have been involved in curriculum development, and activities have been planned so that open conversation is encouraged and developed. Brumfit (1983) recognized that language has shifted from fixed to negotiable within the social context. Not only have students decided what they want to learn, the local populations have determined which languages are taught. According to Brumfit (1995), language learning should be interdisciplinary and instruction should realize its socially relevant status.

The foreign language curriculum has become increasingly socially adaptive, and there has been an increased need for EFL courses in the K-12 schools. Changing communities provide an opportunity for foreign language students to take on a socially reconstructivist curriculum by, for example, helping local citizens of other languages. The local inclusion of various foreign languages has reflected a socially adaptative curriculum in response to the origins and first languages of large populations, e.g., Spanish and Japanese, within the area. Since ACTFL included the goal of learning about foreign cultures in its foreign language curriculum requirements, there has been increased, though often superficial focus on multi-cultural understanding (Ivers, 1990). Through foreign language curriculum, students can develop deeper and broader consciousness, changing the way that they construct their truths.

From the academic curriculum of the Latin Grammar Schools to the social reconstructivist curriculum of schools that adapt to local populations, foreign language goals and methodologies have changed significantly over the past 375 years. Although defined and measured differently, the common goal has always been language proficiency.

The Misalignment of Goals and Assessments

Many foreign language students who learned to read and write a foreign language possess less experience in the aural and oral skills. Since 1981, the American Council for the Teaching of Foreign Languages (ACTFL) has suggested that foreign language instruction include speaking, listening and cultures as well as the traditional reading and writing skills. Although classroom resources have been written according to national standards and promote the instruction of all five skills, all of the skills have not been assessed equally. Stated foreign language goals and the actual assessment of student skills have often been misaligned.

Educators have been required to state their learning goals and objectives daily to provide instructional focus and increase accountability. Lesson plan requirements have often referred to goals as broad statements concerning the intent of the curriculum, while objectives have been more specific, measurable outcomes. For ease of communication in this document, the researcher has used goals and objectives interchangeably to refer to the measurable learning behaviors that emerge as a result of learning.

Assessment has been an important component in curriculum. Designed properly, it can provide meaningful feedback to students, parents, teachers, and communities. Assessment can establish appropriate student course placement, measure student achievement, and provide diagnostic data. It is a valuable step in the cycle of curriculum design. Often considered an ending point, assessment is a pivotal objective in backward curriculum design where it defines the learning objectives. It serves as a measurement of student progress when it is administered at the beginning, during, and at the end of any instructional period (Brown, 1995). Assessment should reflect what is taught in a format that mirrors instruction, and it should focus on what

9

students can, rather than what they cannot, do (Shrum & Glisan, 2005). Theoretically, this is a reasonable proposition.

The creation of assessments can be labor intensive, even when an instructor is simply modifying an existing test. The efficiency of the administration and grading of assessments can vary greatly among content areas and teachers. In some exceptional education classes, for example, where the students have individualized education plans for achievement, assessment is often individualized. This can be a manageable task, given the typical student to teacher ratio. In traditional education classrooms, given the larger number of students, more standardized assessments are administered. Regardless of the content area, assessment should reliably measure student learning according to the stated program goals.

Assessments should be valid, and the measurement given should be reliable. In objective assessment, scores are reliable but the assessment is not always valid. In subjective assessment, the opposite is often true. Authentic assessments that reflect real life applications can be difficult to grade reliably and consistently, making them effective but not necessarily efficient. All educators know that some form of assessment is necessary, but not all educators assess in a way that is aligned with their own program goals.

<u>Research Questions</u>

It was the hypothesis of this study that the stated goals and assessments of foreign language teachers in Florida middle and high schools are not consistently aligned. A random sample from this population was asked to self-report on their instructional practices as they relate

10

to the five foreign language skills as cited by the American Council on the Teaching of Foreign Languages. Following are the research questions used to guide this study:

1. What is the actual allocation of instructional time and assessment on each of ACTFL's five elements of foreign language proficiency?

2. Does the allocation of textbook and computer resources have an impact on the alignment of oral proficiency goals and assessments?

3. Does the level of experience of the teachers, identified as (a) number of years teaching foreign language and (b) level of education, have an impact on the alignment of goals and assessments?

4. Does the amount of target language used in the classroom, by teacher and students, have an impact on the application of oral assessment?

5. If teachers do not include oral assessment in their teaching practice, what is the reason? Hypotheses include (a) lack of time, (b) lack of technological resources, (c) discomfort with technology, (d) test reliability issues, (e) test validity issues and (f) lack of objectivity in assessment.

<u>Organization of the Dissertation</u>

This dissertation is divided into five chapters, a list of references, and appendices. The first chapter includes an introduction and significance of the study. In this chapter, the researcher has provided a brief overview of the history of foreign language curriculum to demonstrate the direction of instruction over time and the results of research on current foreign language acquisition theory. Revealed were some of the gaps in the current research, in terms of foreign

language pedagogy theory versus practice. In the second chapter, the researcher reviews literature that has been published about the existence of a critical period in foreign language acquisition which has an impact on foreign language instruction. In addition, there is a discussion of the importance of assessment of the oral production of language and the difficulties that are associated with it. The chapter also addresses existing research on innovations for oral assessments, the potential complications of technology in the classroom, and current research on foreign language teacher training.

In Chapter 3, the researcher reviews the design of the study, describes the population to be surveyed, and discusses the random sampling process. Chapter 3 also includes information regarding the questionnaire itself. Chapter 4 includes the results of the study and a quantitative analysis of the data collected. Chapter 5 contains a discussion of the findings, their implications for teachers and foreign language program administrators, and suggestions for future research in the area of foreign language instruction. Chapter 5 also explains the limitations of the study and how these factors might be avoided in future research.

CHAPTER 2
REVIEW OF LITERATURE

Introduction to K-12 Foreign Language Education

Since the original publication of its proficiency goals in 1981, ACTFL has recommended

12 years of school instruction in order to achieve second language fluency. Despite this

recommendation, a well-articulated K-12 public school foreign language program is rare.

Foreign Language in Elementary Schools (FLES) programs exist, but suffer during budget crisis.

According to the Center for Applied Linguistics, 25% of public elementary schools had a foreign

language program in 1997. In 2008, there were 15%. For many students, the earliest opportunity

to receive foreign language instruction has been the "wheel" study during the upper elementary

school years. The wheel is an exposure to several foreign languages for equal periods of time

during a semester or school year and is not articulated to the next level in any single language.

This has been ineffective. According to the Center for Applied Linguistics, if the program's goal

is a high level of fluency in the foreign language selected, the program planning committee will

choose the earliest possible start for the study of the language, maximize the time and intensity

of the program at every level, and provide an articulated program that flows across levels without

interruption. Students will be able to continue their study of the language throughout every level

and will have the opportunity to add a second language or change languages at the beginning of

middle or high school. Some students coming from the "wheel" have the opportunity to start the

selected language again in middle school depending upon the availability of foreign language

units (teachers) and the students' reading scores. Students, however, who have not scored

sufficiently on reading assessments are enrolled in intensive reading courses and have no room in schedules for electives such as foreign language.

Two well-articulated and rigorous programs that approximate proficiency in foreign languages are the Advanced Placement (AP) and International Baccalaureate (IB) programs. The AP program is intended for students who have received instruction in a foreign language prior to entry into high school. These two programs are similar in their assessment of each student's reading comprehension, listening comprehension, written and spoken expression in the appropriate language. They differ from other foreign language programs in their inclusion of students. There has been a marked increase over the last several years prior to this study in the number of AP tests taken each year, explainable by wider enrollment (equity), which has threatened curricular rigor (Vail, 2006). Mainstream schooling, which also affects the inclusion of all children into foreign language classrooms, has required an adjustment to the curriculum in foreign language according to McColl (2005). She has suggested that, in order to accommodate the learning needs of such a variety of included students, the curricula be adjusted to address either the "close-and-familiar" or to also include the more traditional, abstract concepts that are required of the more academic curriculum. In 2000, the National Council of State Supervisors for Languages concluded that, in order to be successful in the global community, multilingualism should be basic, and American curriculum should include foreign languages for all students K-12 and beyond. At the time of the present study, despite the "suggested requirements" of foreign language study for at least 12 years, there was no current requirement for the study of foreign language other than to meet a high school graduation requirement or a college admission requirement of two years.

14

When students move to high school, many wish to pursue two years of a foreign language in order to gain admission to a four-year university upon graduation. It is not required for graduation in the State of Florida, only for university admission. For many college graduates, initial foreign language exposure could have been as late as age 16, perhaps too late for language acquisition. The existence of a critical period for foreign language acquisition has been a much-researched topic within the field of linguistics.

Critical Period in Second Language Acquisition

Many state boards of education have invested time and money to create curriculum maps for instruction and adopt standards to which their teachers must adhere and, at times, defend. Regardless of these maps, good teachers know that there is a sequence, from easier to more difficult, in which material should be taught in order to encourage student learning. As children learn, existing schemata are recalled or activated, more elaborate information is presented, and information scaffolding takes place (Anderson, Spiro, & Anderson, 1978).

Human beings learn how to speak in simple language before they can quote literature. Children learn how to add before they learn how to subtract, multiply, and divide. They memorize the periodic table of the elements before they start to mix compounds. In the context of scope and sequence of foreign language curriculum, current practice has not necessarily been supported by research. Many parents and education professionals have assumed that younger children assimilate language more readily than older children and adults. In spite of this belief, American public schools have not offered widespread elementary foreign language instruction. There are several reasons for this imbalance in the elementary curriculum. As English speakers,

15

many do not see a need (at any age) to embark on studies of other languages, since English is one of the most commonly-used languages in the world. Secondly, reading in the base language is considered to be more important than an early start to the acquisition of an additional language. So, additional teaching units have often been allocated to reading. This has not been limited to elementary schools and depends, of course, on student needs as identified by reading scores. In contrast, elementary school-age children have been required to have art and music time. Apparently, attention is paid to research that suggests children can accomplish more academically if they are exposed to the arts at an early age, or the research may not be considered. It may simply be easier to find art and music teachers than foreign language teachers and, thus, creating required planning time for regular classroom teachers.

Foreign language in the elementary years of education has usually been a luxury for those children whose classroom teachers happen to be familiar with a foreign language and can fit some songs or topical vocabulary into the curriculum. If foreign language instruction exists at all, it is often superficial and unarticulated. There has been limited scope and sequence for foreign language at this supposedly optimum age for instruction. The age at which language should be taught has been a topic of debate within the field of foreign language instruction (Curtain & Dahlberg, 2004; Shrum & Glisan, 2005). It is a cognitive, political and practical question for the stakeholders in the nation's schools.

What does the research actually say about the optimum time for foreign language instruction? First, language acquisition is an area of both philosophical and scientific study. In his Cartesian Linguistics, Chomsky (1983) described the universal human nature of linguistic expression, expounding on Descartes' position that humans possess innate, mental structures

whose single purpose is the development of language. Doughty and Long (2003) observed systematic errors in children's first language production and concluded that these errors were caused by inter-language grammar confusion. This was consistent with Chomsky's concept of the existence of a universal grammar.

Second language instruction is warranted in the elementary school curriculum, particularly for speakers of other languages who are learning the local language. What about native speakers of English who want to have a chance in the global community when they have completed their studies? Is there research to support the common assumption that children learn languages easier than adults? Many believe that children are somehow wired to receive additional languages differently than adolescents and adults. Children lose the ability to aurally differentiate certain sounds after age 1, and they can create new linguistic circuits up to puberty (Kuhl,1993.) This has often been explained as children having more brain plasticity. When people of varying ages sustain brain trauma, children are able to speak sooner, after brain trauma to the right hemisphere, than adults, suggesting that adult language is generated in the right hemisphere, and children use both hemispheres (Kuhl, 1993). In studies of age effects on second language acquisition, Silverberg and Samuel (2004) concluded that the system of coding and retrieving second language differs between children and adults. At the elementary school age, a positive correlation exists between memory and language ability (Harley & Hart, 1997) and implicit instruction is effective (DeKeyser, 2000; Norris & Ortega, 2000). These researchers concluded that learners must analyze the language, starting in adolescence, in order to learn it. Foreign language curricula seem to reflect this belief that children and adults learn language differently, necessitating either a naturalistic or instructed curriculum (Doughty & Long, 2003).

In research to investigate the existence of a critical period of second language acquisition,

Johnson and Newport (1989) studied the English language abilities of native speakers of Korean

or Chinese. They found that there was a strong correlation between English language syntax

proficiency and the onset of English language instruction prior to puberty.

Using American Sign Language, Newport (1990) tested whether there was a critical period for

language acquisition. Evaluating the abilities of native signers (first exposure at birth), early

signers (first exposure before age six) and late signers (first exposure after age 12), she

concluded that, even after having used the language for 30 years, the native signers outperformed

the early signers on tests to demonstrate complex syntax (grammar). Consistent with this

outcome, the early signers also outperformed the late signers. These results would suggest a

critical period for language acquisition.

DeKeyser (2000) performed a research study to test the Fundamental Difference

Hypothesis. He built on the research of Johnson and Newport (1989) by using a modified version

of one of their batteries, the Grammaticality Judgment Test. This test provides 200 aural

statements via tape recording relating to grammatical correctness. Subjects provided responses of

correct-incorrect. The Language Learning Aptitude Test, a shorter, oral test, was also

administered in this study. The subjects were 57 native speakers of Hungarian. They were

recruited via Hungarian-American publications and paid a small fee for volunteering to be

participants. The participant group varied in age at arrival, socio-economic status, and level of

education.

DeKeyser (2000) concluded that there was a negative correlation between the age at

arrival and performance on the Grammaticality Judgment Test. High verbal aptitude among older

learners had an impact on level of proficiency achieved in English language but not on the younger learners. The younger learners approached native fluency, whereas only the adults identified as having high verbal aptitude approached this level of second language acquisition. DeKeyser (2000) concluded that children learn language through implicit mechanisms and that adults have lost that ability. Adults acquire a second language, instead, using explicit mechanisms which require structure and problem-solving capabilities.

In response to DeKeyser's (2000) research on a critical period in second language acquisition, Bialystok (2002) claimed that DeKeyser's research did not, in fact, support a critical period in second language acquisition. Bialystok (2002) found contradiction in DeKeyser's (2000) conclusions, including measurements of the older group that did not correlate with those of the younger group, such as linguistically-talented adult learners who scored higher than the adult group norm and an apparent ceiling of the proficiency in the younger learners. Bialystok (2002) seemed to support a critical period in language acquisition but found no support in DeKeyser's (2000) data. Bialystok (1998) had earlier called on the research community to work together in the study of language acquisition. She cited three flaws in research: (a) a definition of proficiency, (b) the relationship between cognition and language ability, and (c) the biological basis for language acquisition.

Flege, Yeni-Komshian and Liu (1999) performed research to evaluate the critical period hypothesis for second language acquisition. The hypothesis of the researchers was that the age of arrival, the frequency of English use, and amount of American education of non-native speakers of English would have an impact on pronunciation of and communicative competence using the

19

second language. The subjects of the research were 240 Koreans who resided in the United States of America. They differed in age upon arrival in the United States.

Flege et al. (1999) utilized three tests. The discontinuity test demonstrated that those who arrived prior to the critical period performed better in the second language. The second test was the pre/post correlation test. Subjects' competence in the second language decreased as the subjects approached the end of the critical, or sensitive, period. A correlation was made between the competence, critical period and age at arrival of the subjects. The third instrument was the matched subgroup. Subjects were selected from the group of 240 and clustered based on variables separate than their age at arrival. Another group was selected based on similar age at arrival. If the latter scores were similar, despite differences in age, it suggested that there was no maturational critical period. The subjects were tested individually in a quiet room by college-aged Korean/English bilingual research assistants. The assistants used a score of 1 to 9 to evaluate the level of accuracy and foreign accent of the subjects. The instrument included 144 sentences, half of which were grammatically correct. The grammatically incorrect equivalent appeared no closer than half a page from the grammatically correct sentence. A native English-speaking, evenly-paced, male, clearly-enunciating voice read the sentences through a tape. The subjects were to specify Yes or No that the sentence was or was not grammatically correct.

The results of the Flege et al. (1999) research were that the later the subject arrived in the United States, the stronger the Korean accent. In a scatter plot, there was a well-defined slope with age at arrival on the X-axis and clarity of accent on the Y-axis. The scores on morphosyntax of the Korean subjects fell within two standard deviations of a control group of native English speakers. This indicated that the effect of the age at arrival had more of an impact on phonology

than on morphosyntax. Age at arrival had more of an impact on accent. Frequency of use of English had more of an impact on accuracy. The amount of American education experienced by the subjects was significant in the rule-based, morphosyntax group. This research suggested that the age at arrival did not have a significant impact on morphosyntax, so it would not appear that there was a maturational critical period for learning a second language.

In another attempt to assess the impact of age of initial acquisition on second language proficiency, Silverberg and Samuel (2004) investigated the ability of learners to move from first to second language. The researchers hoped to determine the extent to which bilinguals' language systems were shared or separate. In reviewing the literature, Silverberg and Samuel identified the findings of eight second language researchers. They found that there were inconsistencies in the reported impact of priming from base language to target language. Silverberg and Samuel hypothesized that much of the conflicting research on language priming from first language to second language could be attributed to the age of the subject at initial second language acquisition. The 72 subjects in the study were native Spanish speakers solicited from posters on a college campus. All subjects learned to speak Spanish first, English second, according to participant surveys. The same surveys allowed participants to identify the age at which they acquired English. The researchers ascertained through proficiency tests that Spanish was, indeed, the stronger language. The group was divided into "early" or "late" acquisition, where seven years of age was the dividing line. The late group was further divided into "high proficient" and "low proficient".

Three instruments were utilized. A language history survey allowed participants to self-rate their skills in listening, speaking, reading and writing in the two languages. This language

history survey was also the basis for placement of each subject. The second instrument was the Boston Naming Test. This was a spiral notebook containing 60 pages of black-and-white images, varying in frequency of daily use (for example, bed vs. abacus). Each image was displayed and each subject was given 10 seconds in which to respond. This exercise was done twice for each subject in English and twice in Spanish. The last instrument was used in a language priming experiment. Subjects were primed with four words that were each similar (in varying ways) to a target word in the other language. Three of these acted to distract the subject from the correct meaning of the target word.

The results of the Silverman and Samuel (1999) research were that the early learners were distracted by semantic (similar meaning) priming. They were distracted, to a lesser extent, by mediated (similar Spanish word, different meaning) priming. The late high proficient subjects were affected by the form (similar English word, different meaning) priming. Most of the results of the late low proficient subjects were not statistically significant. If the priming impacted the subject's ability to access the Spanish word, the researchers concluded that the inhibition was due to linguistic competition, implying that the linguistic sub-structure was being shared. The research of Silverberg and Samuel (1999) suggested that the age at which the second language was acquired had a significant impact on how bilinguals represented and accessed words in their second language.

Current research suggests that there is not a critical period for language instruction and acquisition. However, there appears to be a direct link between the intensity of interaction in the language and the level of achievement in morphosyntax regardless of the subject's age. The research supports a critical period only for phonetic language. Those who learn a second

language after a certain age will generally not be able to acquire native-like pronunciation in the foreign language. Rather than an optimum time for language acquisition, Flege et al. (1999) indicated that the amount of time spent in acquisition was more important. Flege et al. stated that the age of acquisition of a foreign language has a phonetic impact but bears little impact on the ability to communicate in the language. The factor that these researchers found to play a more significant role in language acquisition was the amount of time spent immersed in the language. The research of Silverberg and Samuel (2004) suggested that subjects who acquired a second language at a young age maintained one cognitive linguistic sub-structure in which the languages functioned within one space. Silverberg and Samuel (2004) concluded that the age at which a second language was acquired had a significant impact on how those languages were represented and accessed. The goal was to maximize the frequency and intensity of foreign language exposure.

Regardless of the age at which foreign language instruction begins, researchers have suggested that the intensity of the exposure is the factor that determines fluency. When the recall of knowledge becomes unconscious, implicit memory is engaged (Jacoby & Witherspoon, 1982). Practice is required for the automation of any skill. In the case of language acquisition, automation is fluency. In order for this fluency to be approached, time must be spent on learning regardless of the research relating to the existence or non-existence of a critical period.

DeKeyser (2000), Bialystok (2002) and Newport (1999) claimed that their data support a critical period in language acquisition. Silverberg and Samuel (2004) and Flege et al. (1999) argued that their data support a critical period in language representation and native-like accent

production. Current research supports starting acquisition in the early years so as to nurture at least a phonetic ability.

Further research might identify the cause of the effect of age of acquisition on a person's ability to properly pronounce sounds in a foreign language. In order to teach pronunciation, it would be critical to establish whether the maturational changes are physiological or cognitive. Some subjects appear to have an ability to mimic sounds and can acquire a native-like pronunciation in a foreign language regardless of the age of foreign language acquisition but these are exceptions to the norm. Bialystok and Miller (1999) concluded that the similarity of grammatical structures in subjects' base and target languages must be considered before making conclusions about a critical period. Birdsong (1992) concluded that younger learners can more easily handle grammatical irregularities than older learners. This finding was somewhat similar to those of Silverberg and Samuel (2004) in that the ways of coding and retrieving second language differ between children and adults.

Clearly, language acquisition is a complex human phenomenon. To attempt to prove or disprove the existence of a critical period for language acquisition for all human subjects in all aspects of all languages is perhaps impossible. Researchers have continued to investigate in this field, e.g., examining subjects who are of different ages, manipulating different languages from different home environments to different learning environments, and examining linguistic structures of subjective difficulty. The only true critical period appears to be in accent production, and even then, exceptions to the norm have been noted.

Regardless of the volume of research that has been conducted on the subject of critical period in language acquisition, there has been a noticeable lack of oral proficiency assessment. In

the more common instruments, the language proficiency of the subject has been established by written and aural means. The Modern Language Aptitude Test, administered by DeKeyser (2000), involved aural and written prompts. The Boston Naming Test, administered by Silverberg and Samuel (2004), was a written test. Even the Language History Survey, used by Silverberg and Samuel (2004) to establish subjects' abilities with the four language skills (reading, writing, listening and speaking), has been self-reported by subjects. Flege et al. (1999) measured pronunciation, not speech production, during their 1.5 hour sessions with each of their subjects. Their subjects repeated 21 sentences, after a native-speaker prompt and a pause, which had been previously seen in the same order during a written portion of the instrument. It is surprising to note that there has been a lack of measurement of subjects' authentic oral production of language in attempts to prove or disprove the existence of a critical period in second language acquisition. This was perhaps indicative of the complication that would be added by the numerous associated variables.

Current Standards for Foreign Language Instruction

According to the national standards for foreign languages (Appendix A), there are five language objectives. These are communication, connection, comparison, culture and community (ACTFL). Communication, the central goal, includes the four language skills: speaking, writing, reading and listening. Speaking and writing involve target language production, the more difficult skills, and reading and listening involve recognition. All skills are not equally assessable, however. Traditional multiple-choice tests do not evaluate or encourage higher level thinking or problem solving skills (Im, 2000). More significantly, they do not allow students to

demonstrate language production, only recognition. The most commonly stated foreign language goals involve proficiency in communication using the target language. Communication involves interaction as well as recognition.

Reading comprehension is the most straightforward skill to assess among the reading population. Generally, students read passages and answer questions or select words that logically and reasonably complete sentences. Answers are either correct or they are not. Many speakers of other languages even claim to be able to decipher enough written words to gain limited comprehension in a different target language. Reading comprehension is an attainable and easily assessable skill of keyword recognition.

Listening comprehension can also be easily assessed. Most textbook adoptions include media for aural instruction and assessments and involve an attempted variety of native speakers of the target language. Students' exposure to a variety of accents is important in foreign language acquisition, and even native speaking teachers should employ them. These aural assessments are typically objective with correct, incorrect, or multiple answers. Although they require some special preparation and consideration during administration, they tend to be very efficient to grade.

Writing is a skill of language production. As such, it is a more difficult skill to attain and force students to recall rather than recognize target language. Easy to administer, the inconvenience of written assessment is the grading. Depending on the intent of the assessment, formative, summative or diagnostic, the instructor should provide a certain amount of timely feedback (Sprenger, 2005). In foreign language production, there is also the question of the level of corrective feedback that should be provided relative to the affective filter of the learner.

Written papers of novice and intermediate learners are often rife with mistakes that make the writing difficult to decipher. When providing feedback to such an assessment, the assessor might only correct the most relevant errors in an effort to not stifle the student's motivation. The grading of written assessments is time-consuming but perhaps has as much to do with thoughtful planning and instruction as with assessment.

Speaking is also a skill of language production. It is considered to be the most difficult foreign language skill to demonstrate due to its impromptu nature. Students do not have the reflection time that is afforded them in writing. To further complicate matters, they must often communicate in an exchange which results in an interactive skill. Speaking assessment in the foreign language classroom is often loud, imperfect, difficult to understand, and difficult to evaluate. Even the level and timing of feedback is debatable. Krashen (1982) and Jung (2000) have suggested that too much corrective feedback restricts the desire to produce language by raising the speaker's affective filter. In fact, DeKeyser (1993) found that error correction did not lead to overall performance improvement. Sprenger (2005) and Swain (1995) endorsed immediate feedback to maximize learning gains. A foreign language curriculum expert, Brown (1995) addressed the importance of assessments that measure whether or not program goals have been met. He recognized the four communicative language skills but did not address the practical application of oral assessment. Speaking a foreign language has been considered to be the most difficult skill to attain and to assess, and it has been frequently avoided.

Oral assessment has been receiving much attention from researchers for the past 30 years. School districts and practicing teachers have expressed concern more recently. In Korea, the ministry of education instituted a requirement of oral assessment of English language. In the

United Kingdom, the national standardized test, the General Certificate of Secondary Education (GCSE), has included an oral component for foreign languages. In the United States, the Advanced Placement and International Baccalaureate tests have required students to demonstrate their speaking ability in the foreign language. Although it is a requirement in high stakes testing, teachers struggle to apply oral assessment to classroom instruction (Im, 2000). The foreign language skills include speaking, writing, listening, reading and an awareness of cultures. According to good curriculum design, each of the skills should be assessed. Due to the complex nature of the assessment, oral language is not assessed as frequently as the other skills.

Realizing that there is a need for such assessment but that it is not readily available, the Center for Applied Linguistics (CAL) has contributed significantly to the development of computer-based oral assessment efforts. According to CAL, attaining high levels of proficiency continues to challenge the United States. Depending on the difficulty of the language for English-speakers, CAL has estimated that the average language learner requires 720 hours of instruction to achieve an advanced proficiency rating. For a language that uses different characters, such as Russian, CAL has estimated 1320 hours. Most foreign language students pursue their minimum requirement of two years, which amounts to 180 hours. CAL has claimed that it will be an urgent situation if effort and resources are not invested into the development of a language-proficient society that includes high levels of proficiency.

Populating upper level classes in foreign languages has been a program goal and continuing issue in high schools. Stewart-Strobelt and Chen (2003) conducted a survey focused on student considerations when deciding to study foreign language. The purpose of the study was to illuminate changes that foreign language teachers might consider in an effort to increase upper

level enrollment. Most upper level students stated that they continued foreign language study because of an interest in the language and its associated cultures. Hernandez (2006) referred to this positive attitude toward the language, its native-speakers, and their culture as integrative motivation. In contrast, those who were instrumentally motivated to study a language in order to secure a position or achieve some specific objective have not continued to as advanced a level. His recommendation for motivation to the upper levels was for instructors to focus on a variety of activities that promote integrative motivation. Verkler (2008) also suggested a variety of context-rich activities that make language "come alive," in order to retain students through the upper levels. In her survey of upper level Spanish students, students identified integrative factors first, followed by instrumental factors, as their motivation to continue their foreign language study. Interacting in the foreign language in a real-life sense, through speaking and listening, is instrumental.

Perhaps speaking a foreign language is not perceived by all learners to be a necessary skill. Indeed, it might not be a future requirement for language learners, hardly worth all the effort. It is interesting to examine the reasons that people chose to study foreign languages, as well as their motivations to continue to the advanced levels, toward proficiency.

<center>Motivation of Foreign Language Learners</center>

In any classroom, the students present can have many different experiences. Bang (2000) identified five affective variables that influence Korean students' ability to learn English: beliefs about language learning, motivations, learning styles, personalities and affective states. Lightbrown and Spada (2006) contended that differences among learners can be attributed to

<center>29</center>

motivation, intellectual ability, personality and learning preferences compared to methods of instruction. Oxford and Ehrman (1993) identified motivation, anxiety, self-esteem, tolerance of ambiguity, risk-taking, language learning style and age as variables that affect student ability to acquire foreign language. They claimed that understanding the role of these variables might help teachers vary their methodologies. In 2003, Ehrman, Leaver and Oxford discussed some differences in learning styles, learning strategies and affective factors. They claimed that intrinsic motivation was the extent to which someone learns a language because of a desire to do so and the satisfaction gained from doing so, an expectancy-value model of motivation for individual self-efficacy.

In a unique case-study, Samimy (2008) interviewed a non-native speaker of Arabic who had achieved the "superior" rating on the ACTFL Oral Proficiency Interview, a nearly impossible feat for a non-native speaker with limited native-speaker exposure in a less commonly taught language. His ability was also measured by two college professors, to ensure reliability of the superior rating. The subject was highly motivated to learn the language and had actually immersed himself with headphones and audio programs on his personal stereo, creating for himself an "Arabic bubble." This individual was highly motivated to learn his selected language, one that was not offered to him during his K-12 schooling. He required all language skills and relied heavily on aural comprehension for immersion.

According to research conducted by Williams, Burden and Lanvers (2001), the motivation to study foreign language declined with age, sex and language choice. They concluded that boys were less likely to continue in foreign language than girls and that they were more likely to study German than French. In a more probing, qualitative study, they found that

French was considered by boys of 12-15 years of age to be a feminine language, that boys were less likely to engage in "tedious study" or to be seen as trying too hard to pronounce it properly. They referred to cognitive psychology expectancy-value models for explanations of motivation, or lack thereof, to continue in foreign language. Expectancy-value models have purported that learners weigh their likelihood of success against their perceived value that is attached to such success. The English Office of National Statistics has been quoted (1999) as stating that girls far outperform boys on the GSCE (General Certificate of Secondary Education) in modern languages of French, German, and Spanish. Williams et al. (2002) surmised that one reason in the complex equation was that foreign language is perceived to be a female-dominated subject due to the lack of male teachers in the field and the female bias reflected in course syllabi.

Pritchard and Loulidi (1994) claimed that in Northern Ireland, there were significant differences between the Catholic and Protestant school curricula based on belief and funding that cause languages, history, and the arts to be more central to the former and math and sciences to the latter. They identified significant differences between the genders which were much more favorable among girls in terms of the attitude toward the study of languages. Boys had reportedly more favorable attitudes toward German than French, similar to the findings of Williams et al. (2001). German is described by the quoted students as being "strict," "guttural" and "easy to pronounce."

In another UK study, Dewaele (2005) indicated that for Dutch students English was a preferred subject over French as a foreign language despite the greater length of required study in the latter language, it being the second language to Flemish-speakers in the Netherlands. He observed that there is societal-linguistic tension due to neighboring Belgium and its complex

linguistic boundaries. He discussed personality issues, such as extroversion and introversion, as they relate to second language versus foreign language study. In this case, French was the second language and English was the foreign language with less opportunity for authentic exposure and increased foreign language study anxiety.

Every foreign language learner does not require training in all five skills. Scholars might only require the ability to decipher written language in order to read research. Military radio operators might only require the ability to decipher spoken language, and perhaps some cultural knowledge, to better understand the language as it is used by native speakers. For most other language learners, the need to communicate is interactive. In addition to listening, the learner will ultimately have to speak. Contrary to common classroom practice, most language learners will not have to write other than to facilitate the visual component of the learning process. Reading and writing should be used as tools and not as ends in themselves (Curtain & Dahlberg, 2004). Most individuals study foreign language in order to communicate with native speakers of the language.

<u>The Importance of Oral Assessment</u>

According to Shrum & Glisan (2005), assessment is an essential part of teaching and should reflect what is taught. Successful assessment improves learning, instruction and program effectiveness. To ensure relevance, authentic, real-world assessment has been recommended. Instruction and assessment should be directly connected. In 1989, Clifford stated that the end objectives of foreign language instruction had been inadequately defined and communicated. Some of the earlier oral assessment practices were diagnostic assessments within the field of

bilingual education, such as the Spanish-English dominance test created to stream 6- and 7-year-olds into the appropriate class, the Navajo-English Dominance Interview, and the Oral Placement Test for adult speakers of English as a Second Language. The assessments were close to the practical situation for which each was created, and they were intended for use with simple materials and minimally-trained testers (Spolsky, Murphy, Holm, & Ferrel, 1972.).

Over the last 30 years, there has been an intent in educational reform to include oral assessment in foreign language curricula. In Scotland, educational reformers called for (a) a diagnostic oral assessment at least once during each student's required foreign language study, (b) that orals be included in the public exam at age 16 (O-Levels) or 18 (A-Levels) for future employers or universities, and (c) that there be a sampling of oral assessment to measure program effectiveness (Assessment of Standards, 1977). In Wales, there has been political and cultural pressure to require instruction of Welsh, offered as either a first or second language, starting in the primary schools (Price, 1980.) Public information is written bilingually, and aural and oral proficiency are critical to Welsh language use. Many languages, however, have not had a written tradition. Marshall & Peters (1989) explained that, despite being a language whose oral traditions indicate Maori culture, only 15% of the Maori National Exam assessed oral proficiency. As in Wales, it became a political and cultural situation. The State Department created a project to investigate the feasibility and reliability of teacher-based assessment, using a style of assessment that was culturally appropriate to the Maori, in other words, oral. In more recent years, even countries with cultural resistance to oral expression have moved towards instruction in oral proficiency. The Hong Kong Examinations and Assessment Authority has progressed from norm-referenced to standards-referenced evaluation measured during classroom

activities over a two-year period in order to pass the Hong Kong Certificate of Education Exam (Davison, 2007).

Oral proficiency includes many skills such as appropriate use of grammar and vocabulary, the application of appropriate sentence structure, and a certain rate of fluency. Chambers (1997) defined the commonly misused term "fluency" as including the number of pauses, the length of speaking between pauses, and the location of pauses (during or after sentences) relative to speakers' patterns of pausing in their first language. Oral proficiency also includes pronunciation. Communication has been the goal for speakers of foreign languages, and yet communication has often not been achieved because the listener does not understand what the speaker is trying to say (McCarthy, 1978). In many languages, the transfer from written words to spoken words is not automatic, especially in non-phonetic languages like French and English. Farber (1991) suggested that learners mimic native speakers to achieve a comprehensible accent. He even proposed that language learners use a form of transliteration, the creation of a writing system that is phonetic, to assist in learning to speak the language being learned.

Practically speaking, oral proficiency is the most utilizable skill in Western cultures. Many adults testify to linguists that they have taken so many years of a foreign language and yet cannot speak a word. Most foreign language students will never write a word outside of the classroom. In 30 years of French language use, the researcher has never written a word in French while in a French-speaking country! Oral proficiency is at least as practical for most students as the other three skills; thus, it should be properly integrated into the curriculum. As such, it should be assessed, in addition to the more efficiently-tested reading, writing and listening skills.

According to Benjamin Rifkin (1996), when teachers do not assess students' oral language production, it lessens students' perceptions of the value of speaking in the target language. It is a feared, though necessary, skill.

Oral assessments should measure different competencies than written assessments. Hughes (2004) argued that grammar and a sophisticated and varied vocabulary that is favored on advanced proficiency tests, such as advanced placement tests, are not used by most native speakers and that this is not a realistic or authentic assessment of spoken language. The visible bias of which she speaks is the written bias of language. Oral competencies are particular and should include linguistic competencies, such as circumlocution, sociolinguistics, discourse management, and fluency in addition to the written production skills of lexis and grammar (Hughes, 2004; Paulus, 1998.) Other stated competencies include interaction, building topics, turn-taking, pronunciation, fluency, grammatical ability, and socio-cultural ability (Boyles, 1994; Gonzalez-Pino, 1989; Kramsch, 1986; Magnan, 1991; Nunan, 1989). In spite of standard lists of stated competencies that must be achieved, oral language development is multidimensional and discontinuous. It is fluid and involves a combination of knowledge and skills (Young, 1995). Clearly, these skills, identified by the experts in the field, are very different than the other communicative goals in foreign language. The stated foreign language communicative goals in the United States were created by ACTFL.

ACTFL has served as the national governing body for all foreign language curricula, and foreign language K-12 standards are based on ACTFL proficiency guidelines. The ACTFL proficiency guidelines are based on a rating scale and assessment procedure developed in the 1950s by the Foreign Service Institute of the U.S. Department of State. Government and

academic groups provided grant money to ACTFL and the Educational Testing Service (ETS) in the early 1980s, to develop and publish these guidelines. In 1999, ACTFL revised the guidelines to divide the oral proficiency rating scale into additional sub-levels that would provide a more accurate rating and tangible evidence of individual progress. This is the only change that has been made, and it was made in response to criticism that the "intermediate" rating was too broad. The purpose of ACTFL's proficiency guidelines has been to assess a subject's general language proficiency against a scale of exponentially ascending proficiency levels. The linguistic tasks are designed to increase in complexity from level to level as the ACTFL-trained interviewer skillfully probes the subject. The Oral Proficiency Interview (OPI) is the gold standard in oral assessment. It was intended to be used as a diagnostic test, one that would establish someone's proficiency in a broad and general sense, rather than as a periodic assessment (Brown, 1995; Clifford, 2003.) It was not intended to be used as a performance achievement tool, but institutions, instructors and programs have modified it many times to create an interview that suits their needs.

According to their website, ACTFL currently offers OPIs in Arabic, Chinese, Dutch, English, French, German, Hebrew, Hindi, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Ukrainian, Vietnamese, and others. The Full OPI Tester Training Workshop is four days, costs $850 for non-ACTFL members, $685 for members, requires that the trainee hold a bachelor's degree in the language in which rater certification is sought, and a personal oral proficiency rating of "advanced" or higher. Successful full tester training results in full certification. A Modified OPI Tester Training Workshop is shorter, costs less, requires the lower personal oral proficiency rating of "intermediate high," and results in limited certification.

Trainings are offered to individuals at ACTFL training locations, and to groups at institutions as requested. Outside of Florida, some K-12 school systems currently require that foreign language teachers be trained as OPI raters in order to maintain teaching certification.

Test validity and inter-rater reliability are common research topics by ACTFL's testing branch, Language Testing International as well as language researchers. Surface and Dierdoff (2003) found high reliability among 5,881 oral proficiency interviews in 19 languages, especially among experienced raters, attributable to well-articulated rating criteria, effective training and an effective interview process. O'Loughlin (2002) studied the effect of gender on oral proficiency ratings and found that results were reliable regardless of the gender of the rater and the interview subject. Henning (1992) compared assessments given to subjects by ACTFL-certified OPI raters and naïve native speakers, and found that they concurred on ratings using the ACTFL guidelines. Henning suggested that the guidelines be used on tests. Despite the intention of the OPI rating as a diagnostic tool, Meredith (1990) demonstrated how it can be modified and successfully applied as an evaluative tool, finding that it correlated to the subjects' number of years of study.

There have been many alternatives to the ACTFL OPI, which are referred to as Modified Oral Proficiency Interviews (MOPIs) or Simulated Oral Proficiency Interviews (SOPIs.) These are often created as collaborations between ACTFL and another institution, where face-to-face interviews are not practical. To test validity, ratings are compared between an OPI and the SOPI for the same subjects (Cafarella, 1997; Clark & Hooshmand, 1992; Ke, 1993; Kenyon & Tshirner, 2000; Kuo & Jiang, 1997; Stansfield & Kenyon, 1992, 1993; Stansfield et al., 1990.) Subject preferences for face-to-face interview versus recorded interview seemed to vary, both culturally-linguistically and individually. There is a Hausa Speaking Test, a Chinese Speaking

Test, a Portuguese Speaking Test, an Italian Oral Proficiency Test, an Indonesian Speaking Test, a videoconferencing interview test of Arabic and Russian, and a German Speaking Test. Many of these languages have been known as less commonly taught languages (LCTLs,) so the creation of modified oral proficiency interviews was practical.

There have been many interview techniques in addition to the ACTFL OPI and SOPI. In 1979, Rosser created an oral proficiency interview whose scale included rater consideration of subject fluency, pronunciation, grammar use, vocabulary use, and comprehension of questions. Subjects were rated according to minimum and maximum acceptability for contextual professional situations. Yorozuya and Oller (1980) created a 10-point scale that would measure subjects' grammar, vocabulary, pronunciation, and fluency during an interview.

Some foreign language oral proficiency evaluations are part-interview, or are not interviews at all. Traphagan (1997) administered an oral interview that included a warm-up, a grammaticality judgment test, a picture description, a focus question, and several open-ended questions. Open-ended questions were necessary in rating younger children because whereas the subjects' language production was limited by lower cognition, they were interested in providing emotional talk. Verkler (1994) applied a simple oral assessment that included three images. Students were to select the image of their choice, think about it for 1 minute, and speak about it for 30 seconds.

In foreign language research, researchers have created, modified, or applied existing oral proficiency tests. Foreign language classrooms can do likewise. Ricardo-Osario (2008) found a combination of performance-based and traditional (pen and paper) assessments in undergraduate foreign language education classes. The conclusion was that assessment procedures for oral

proficiency were inconsistent. Communicative competence is established by a learner's ability to communicate in the target language in real situations, but Ricardo-Osario (2008) found that translation superseded oral interviews. Even in a university Spanish conversation class, the final assessment was written!

The Difficulties Associated with Oral Assessment

In general, most students dislike assessment. Making it even more difficult, anxiety has been shown to be more related to speaking than any other foreign language skill (Horwitz, Horwitz, & Lope, 1986.) Attempting to speak a foreign language involves risk-taking, vulnerability, and intuition. The ability to communicate orally is also strategic, involving social affective strategies, fluency-oriented strategies and negotiation of meaning in an interview (Natatani, 2006). Testing for oral proficiency is particularly difficult, especially for novices. Anxiety can have an impact on any performance. In oral proficiency testing of interpreters, Choi (2006) found that novice interpreters experienced more interference from errors that they made than more advanced interpreters who corrected and continued. Choi (2006) suggested that novice interpreters should have a different scale than those with more experience, in order to achieve a more accurate rating for novice speech samples. Oral proficiency is often the most feared assessment for students. It can also be the most difficult assessment for teachers to administer and evaluate, due to its complexity.

Oral evaluation can be time-consuming depending upon the length of the test, the number of prompts, and the nature of the speech sample being gathered. In some oral proficiency of second language research, interviews consumed more than two hours per subject. Hurman (1993)

noticed a range of instructor perspectives on oral evaluations and scoring techniques applied to oral interviews. The interviews consisted of discussion of stimulus material, discussion of a personal interest topic, and conversing on a prepared topic. Conversing on a prepared topic is typical in oral proficiency competition known as impromptu or extemporaneous speaking. Hurman found that conversing on a prepared topic was the least reliable of the assessments. Speaking on a prepared topic is not necessarily an oral proficiency speech sample but can be evidence of good preparation. Hurman found that some raters graded more severely because they had higher expectations since the subjects had had the opportunity to rehearse, ahead of the actual assessment. To circumvent this, Hurman suggested that raters interrupt the speaker during such an oral assessment to ensure the speaker's ability to converse on a prepared topic.

When class time and individual requirements are incongruous, computer assistance has often been suggested. Despite advances in computer science, however, it has been difficult to automate language proficiency assessment in an effort to increase efficiency. Software exists that will count the number of words, but the number of words used, with no attention paid to the finer points of speaking that achieve communication, is not necessarily oral proficiency. The Center for Applied Linguistics has assisted organizations and universities, both internationally and within the United States, with the creation of oral assessment technology. This has not been commonly used by K-12 school systems.

Many instructors and students are uncomfortable with technology. Burnett (1998) found that, in spite of the teacher's motivation to use computer-assisted language learning, technical problems often impeded the ability to sustain lessons. Additionally, it was noted that the activities often favored computer literacy over linguistic proficiency. Technology should be the

means not the end. When Cummings (2005) surveyed K-12 Spanish teachers, she found that despite increasing pressure to incorporate computers into K-12 classrooms, many teachers were using them only for routine classroom administration such as grade-keeping and attendance. She found that many teachers were resistant to using computers for instruction and assessment due to unfamiliarity with technological applications, no time to learn, and intimidation by student knowledge.

When K-12 foreign language instructors have access to reliable language computer laboratories, the process of data voice gathering can be more efficient than the cumbersome one-at-a-time oral assessments of the past. Several types of software exist, such as Sanako, which allow simultaneous voice gathering in a language computer laboratory for later playback and evaluation by teachers. This removes the classroom management issues that can develop while assessing oral proficiency one at a time, but it does not answer the question of efficiency since the instructor must listen to individual sets of voice data at a later time.

Technology exists by which algorithms can be coded and voices can be compared (South, Rugelbak et al, 1993), but that measures speech, not oral proficiency, to a certain standard. Turner (2000) addressed rating scales for the voices that have been used on performance assessments. Such devices only measure accent production and rate of speech, not communicative language. Furthermore, such software is not sophisticated enough to accommodate the variety of comprehensible accents even among native speakers.

Similar to the efforts to digitally measure oral production accuracy, Kozaki (2004) created a program that would accept multiple correct answers in order to assess medical translators. A higher level of proficiency was certainly required in translation, but in this case it

was limited to one context. It was not an oral assessment as in the K-12 foreign language curriculum. It did not measure proficiency since it involves translation rather than free conversation.

In digitized word lists, James, Bowsher, and Simpson (1991) found that the effects of band width, resolution, and decibel level caused the speech intelligibility to be varied among six male and six female subjects in 12 different digital sound conditions. The researchers concluded that the computerized measurement of voice data language was not reliable at lower levels. The sound settings can have a detrimental effect on oral assessment reliability.

Reliability has often been called into question on live oral assessments as well. There are many factors that can influence the assessor's reliability. Jafarpur (1996) determined that there was a significant difference in evaluations performed by native speakers and non-native speakers of the target language being assessed. Jenkins and Parra (2003) found that non-verbal behavior of subjects during an interview affected the rating given. They compared eight international university teaching assistants, four native-Spanish-speakers, and four native-Chinese-speakers, rated by the same raters, and concluded that the subjects' behavior created a perception of linguistic competence. They suggested that, in order to maintain test reliability, the rater should modify the power dynamic to reduce interview asymmetry. Davison (2004) found that there was considerable variability among teacher ratings of written English language by Cantonese-speaking students. According to the outcomes of this study of teacher assessments, teachers could be clearly divided into two groups of raters--those who follow a checklist and those who rely on their own professional experience and judgment. Davison concluded that in foreign language holistic assessment, one should rely more heavily on the professional judgment of

teachers themselves. Ross and Berwick (1992) listened to recordings of 60 OPIs and found that

there was variety to the amount of interviewer-subject interaction. They noted that some

interviews contained qualities of conversations and suggested that the extent of accommodations

made by interviewers might be considered when rating oral proficiency of the subject. Kasper

and Ross (2007) found that the interviewer's management of the subject's comprehension of

interview questions was critical to test reliability. They used the term vertical questioning to refer

to the interviewer's next question being based on the student's previous response. They used the

term, horizontal questioning, to refer to the interviewer's next question being based on

anticipation of an environment where there was a probability that the subject would not

comprehend the question. They found that the situation of the appropriate reactive (vertical) or

proactive (horizontal) question had an effect on reliability. The training of evaluators has been

critical to the reliability of assessment (Purschel, 1981) and the avoidance of bias (Lumley &

McNamara, 1995). Such training takes a commitment of time and money.

Reliability has been critical in standardized foreign language tests such as the College

Board's Advanced Placement tests, the National Board of Professional Teaching Standards

proficiency tests, and the United Kingdom's Ordinary Levels (O-Levels) and Advanced Levels

(A-Levels). Rater assessment requires special consideration, whether the voice is live, as in an

oral proficiency interview, or a voice-recording, as is the case with the oral component of

Advanced Placement tests. Special care is taken to retrain assessors at regular intervals in an

attempt to standardize the raters' assessments. In an effort to standardize each student's rating,

multiple raters have often been placed to perform the grading of orals on high stakes tests. Where

there has been significant inter-rater difference on any assessment, more analysis should be

performed, often involving additional raters. It has been, however, an expensive process in terms of the time involved in each assessment as well as assessors' travel. Test rating, or "reading," for high stakes tests has normally been performed at a central location for the purpose of training and having multiple assessors available. There are high stakes tests that have no oral component, presumably due to the expense of oral assessment. As one example, the College Board's Scholastic Aptitude Tests in foreign languages have no oral assessment. Brantmeier (2006) noted that foreign language placement tests at the university level have increasingly become Computer-Based Tests (CBTs.) CBTs involve demonstration of reading comprehension, vocabulary knowledge, grammar use, listening comprehension and, in some cases, culture. Due to the presumed difficulty of gathering and rating the speech sample, there has been an absence of oral assessment for placement testing.

Inter-rater reliability is of central concern on any standardized oral proficiency test, particularly since many standardized tests carry high stakes. Reliability is a consistently-examined component in ACTFL's Oral Proficiency Interview not only by ACTFL but also by foreign language and linguistics researchers (Ross, 2007; Salaberry, 2000; Thompson, 1995.) Salaberry (2000) found that the OPI was neither valid nor reliable despite changes in 1999. Thompson (1995) found that the OPI raters were reliable but that second raters were three times more likely to score subjects lower when first rater discrepancies were noted. In his study of OPI ratings in English as a Second Language, French, German, Russian and Spanish, Thompson (1995) found that the five languages had different patterns of agreement among raters. He also noted inter-rater reliability was called into question at some levels of proficiency more than others. Ross (2007) researched the reliability of the OPI by testing the same candidate by

different interviewers three months apart. He found that the candidate slipped down one level

after three months, according to the first rating, although the second rating of each interview was

consistent. Ross examined interview style as a variable and found that, despite the first

interviewer style and tone and the subsequent interviewee footing, the second rating caused the

overall rating to be consistent. Ross found that differences in interviewer style can lead to varied

ratings on the OPI.

The OPI validity has also been called into question (Johnson, 2000; Mikhailova, 2007).

Johnson (2000) took exception to the Educational Testing Service (ETS) statement on test

validity that the OPI tests students' speaking ability in the context of a real life conversation

(ETS, 1989). Rather, Johnson claimed that the OPI was a survey research interview-response

type evaluation, not real-life conversation. Mikhailova (2007) challenged the validity of Russian

ratings of Intermediate High, and above. She claimed that the non-elicitation of the description

category may cause unreliable ratings for "advanced" and "superior" on students.

Native speakers have sometimes been tested as a basis for test validation. Native speaker

performance cannot serve as a criterion for non-native speaker validation, however, because the

needs and challenges of each group differ (Jafarpur, 1996.) Ellis (1984) tested middle-school

aged children, who were both native and non-native speakers, with a focus on communicative

strategies. The observation was that, although the native speakers used more language, both

groups of children made errors.

Rating scales that are used by the raters have also been challenged in terms of test

reliability. Bruhn (1989) found fault with the OPI rating scale, claiming that (a) native speakers

were not consulted in its creation, and that (b) the OPI rating of "native speaker" was vague.

Bruhn compared three standardized tests: the Foreign Institute Test, the OPI, and the Summer Institute of Linguistics Bilingualism Project in Senegal. The findings were that each test used different sampling techniques, testing formats, evaluation techniques, and goals. The conclusion was that it was impossible to develop a universal rating scale.

Davison (2004) conducted research on teacher scoring in response to claims that teachers assess inadequately especially in high stakes English tests by Australian teachers in Hong Kong. Some teacher-raters followed a set check-list, and some teacher-raters relied on their own professional judgment. The conclusion was that traditional notions of validity should be reexamined in favor of the professional judgment of teachers in the field.

In the United Kingdom, the oral assessment criteria of the standardized General Certificate of Secondary Education (GCSE) were revised so that the test, especially the free-conversation test, would conform to National Curriculum Standards. Chambers and Richards (1992) found that there was a lack of consensus on what is expected of language learners and how to grade them. They examined the reliability of teacher assessments of the free-conversation component of the GCSE in 1993 and found that the orals were not organized consistently among the native-speaker and non-native-speaker, teacher-graders. The teachers reported that they would grade more reliably if they were able to select the test component that they would grade based on personal preference rather than it being assigned to them based on their native language. In 1995, Chambers and Richards (1992) again analyzed the free-conversation portion of the GCSE. They found that the free-conversation was a valid testing instrument. Unlike an interview, the tester did not do most of the talking. Rather, there was more natural interaction between the subject and the rater who was an unfamiliar teacher. However, they found that the

interaction had problems of authenticity, that there is a lack of expertise of teacher-raters, a lack of reliability in the grading of the free-conversation, even a lack of clear marking criteria with terms such as "linguistic quality" and "complexity" not clearly-defined. There was grading disparity between raters on structure complexity, fluency, and idiom use.

It is difficult to find absolutes in such a subjective assessment as language proficiency. Oloafe (1992) proposed a hierarchical assessment model that considered the communicative value of what the subject was saying. Appropriate and situational vocabulary application would be more valuable than lexico-grammatical structures according to this scale. Kugler & Euerle (1992) created an assessment that involved the objective assessment of subjects' monologues and dialogues scored as a ratio of errors to total words spoken. Luecht (2003) proposed that there were many models and testing considerations. He proposed a multistage complexity model that included theoretical considerations, e.g., language phonology, structure, context, or holistic competency. Secondly, it included test development considerations, e.g., task demands, challenge factors, and complexity. Lastly, his model included psychometric considerations of how the data would be collected and scored, e.g., individual items, according to a rubric, against a scale. Luecht (2003) concluded that only multidimensional oral proficiency assessment designs with aligned theoretical constructs, test design, and psychometric scoring would be reliable and valid. A multifaceted Rasch analysis of rater outcomes, such as bias analysis, can help develop a standard-setting procedure for raters (Kozaki, 1995; Lumley & McNamara, 1995.) Chaloub-Deville (1995), using a multidimensional scaling tool, found that when native-speaker raters were trained to grade subjects' speech samples holistically, the raters were not consistent. The recommendation was that a generic component scale should be reconsidered. Language is

intended for specific purposes, so test content, as well as the criteria for judgment of language learner performance should also be specific (Douglas, 2001).

There are many scales that can be adapted by the classroom teacher ranging from the simple to the complex. Additionally, there are many ways to simplify the task of oral assessment in order to make it more manageable for instructors. There is a vast body of research that offers suggestions for K-university foreign language instruction both on pedagogy and assessment. Reading or collaborating with colleagues might help even those who lack creativity. Foreign language teachers have been urged to pay closer attention to the curriculum standards and align all, not just those easy to assess, of their program goals and assessments.

<center>Suggestions for Incorporating Oral Assessment</center>

The American Council for the Teaching of Foreign Languages (ACTFL) created the Oral Proficiency Interview (OPI), which was a series of questions posed by a trained interviewer. Although OPI was intended as a diagnostic tool, an interview can be a means by which to assess achievement. Kim (1980) recommended a general process for developing interview oral assessment that included setting the interview test objectives, defining content and controlling the types of questions posed, observing and recording the responses, and evaluating the results. Henning (1983) compared three oral assessment formats (interview, imitation and sentence completion) and found that imitation was most effective for pronunciation. When fluency, grammar, and pronunciation were combined, however, the interview provided the most valid results. Higgs (1987) proposed that oral proficiency tests should be individualized, interactive, and flexible and that they should test the communicative function of learners. Rifkin (1996)

<center>48</center>

created a simple five-point scale with which to measure daily oral assessments. The scale included communication, content and originality, grammatical accuracy, lexicon, syntax, and fluency. Although intended as an ESL placement tool, Lazaraton and Riggenback (1990) examined the Rhetorical Task Examination (RTE), a test that incorporated two scales: one to measure participant completion of practical tasks and one to measure participant ability to accomplish linguistic tasks. They concluded that the combined scale of the RTE had future implications for oral assessment. With the increased focus on oral proficiency, there have been many practical suggestions for oral assessment both via a modified interview technique and other ways of gathering and evaluating speech samples.

Norrish (1978) proposed monthly oral assessments that were administered in combination with aural prompts from tapes, and increasing in linguistic demand. Oral assessment should start during the first year of language instruction rather than the second or third year (Frink, 1982; Norrish, 1978). Nibungo and Williams (1996) suggested that during early acquisition, scores be based on vowel and consonant pronunciation, rhythm and clarity of speech, and expressiveness. Intermediate and advanced students would be evaluated via formal and informal presentations, intelligibility, fluency, content, originality, and spontaneity.

According to Curtain and Dahlberg (2004), students should be given the opportunity to perform oral language in communicative ways. Small group, varied, actual-life oral assessments are effective and efficient when organized properly (Robison,1992.) In order to be valid, assessments must be defined for the students and must correlate to what was taught. Mentz (2002) suggested small-group projects that involve oral presentations. Gan (2008) studied peer group oral discussions under assessment conditions and found that students applied collaborative

reasoning, negotiated and co-constructed discussion formats, and that the discussion simulated normal speech patterns.

Szostek (1994) observed that students receive little engagement in target language outside of the classroom, so teachers must create such an environment inside the classroom. In this research, Spanish 2 students were engaged in cooperative learning via interviews, team learning, group investigations, and think-pair-shares. The research of García & Hernández (2007) supported the use of Information gap activities, to increase student exposure to second language, to reduce speaking anxiety, to increase cultural awareness, and move students toward linguistic automaticity. Information gap activities involve student pairs, each with visual access to different pieces of information, forcing students to negotiate orally and aurally in order to complete the task. Ikeda (1998) found that when students were paired as alternating interviewer and interviewee, there was reduced stress, authentic participation, less speech asymmetry than in a traditional, one-sided interview. There was also an increase in student participation and engagement. O'Sullivan (2002) researched the impact of pair-task performance on students' familiarity with partners. There was an acquaintanceship effect which indicated that those who knew each other scored better on paired task performances. Lazaraton and Davis (2008) examined the effect on paired assessment scores of gender, first language, and proficiency of students paired. They found that during paired-format oral assessments, the proficiency demonstrated was fluid, and was dependent upon the language-learning identity that both interlocutors bring to the task. Swain (2001) found that student dialogues were valuable not only for present student oral assessment. They were also valuable for the development of future targets for measurement based on teacher observation of their students' linguistic negotiations

and the inferences that they can make about them. Othman and Poh (2005) concluded that, when presented with meaning-focused opportunities to negotiate meaning, cooperative learning had a positive impact on student performance for oral assessment.

McNamara (1997) found that student interaction through interviews and role-plays broadened the social dimensions of language acquisition in addition to the cognitive and psycholinguistic demands. Berrier (1989) compared interview and role-play as assessments and found that interview was more impersonal and tended to be assessed on grammar and linguistic proficiency. Berrier concluded that role play was a more authentic assessment, since it required increased sociolinguistic skill, and tended to be based on achievement of communication. Oral assessments should be as authentic as possible (Clark, 1981; Hoekje & Linnell, 1994; Im, 2000.) Troyan (2008) examined a co-curricular project-based assessment on world-wide water use, incorporating student knowledge of science and humanities. Student language use was interpretational and presentational, and student performance was assessed via role play, journalism articles and public awareness posters. Students were expected to meet various assessment criteria throughout the project. Once they met or exceeded each particular expectation, the student group was allowed to move to the next.

Van Eunen (2002) discussed oral assessment via a German-language-only living space located in a separate location on a high school campus. Students were expected to use only German language as they accomplished real-life simulated communicative tasks and were assessed using a point system. Van Eunen found that real-life simulation had high memory value among the students. A similar "city" would transfer to any oral language.

Shrum and Glisan (2005) suggested rubrics, portfolios, peer- and self-assessments, projects, and probes to measure oral proficiency. Mathews & Hanson (2004) found that the requirements of a student portfolio of student written work plus an oral proficiency test, were instructionally sound and practically feasible for classroom foreign language teachers. Padilla, Aninao & Sung (1996) found similar outcomes and recommended the requirement of an annual portfolio for each student for each year of K-12 foreign language instruction. This would serve as evidence of progress toward proficiency. Periodic oral assessments could also be included in portfolios and give students a sense of their own oral proficiency achievement. Mir (2006) examined oral assessment through journals. Students recorded 10-15 minute personal monologues, exchanged them with their peers, and listened and responded to each others' oral journals. It was concluded that this task increased student motivation, self-confidence, fluency, awareness of and focus on the oral skills on which they were assessed.

In her self-assessment research, Saint Léger (2009) found that student affective filter was lowered and student motivation was high. Activities could be easily varied and were perceived by students to be relevant. During a 12-week period, students set their own goals and self-assessed at four weeks, six weeks and 12 weeks. St. Léger found that student fluency, use of vocabulary, and confidence increased and that the application of self-assessment benefited students on cognitive and affective levels. McMillan and Hearn (2008) asserted that self-assessment empowers students to guide their own learning and that correctly implemented, students monitor and evaluate their own learning and identify strategies that will cause learning gains. With mastery-type goal attainment, learning becomes more meaningful to them. Through self-evaluation, they learn to discern between current and desired outcomes, raising their

metacognition and self-efficacy. With self-assessment, foreign language instruction moves from assessment towards motivating and enhancing learning. Self assessment, according to McMillan and Hearn (2008), is too subjective to be considered effective for formative and high stakes testing.

Hughes suggested in 2004 a five-point grading scale that measured accent, grammar, vocabulary, fluency, and comprehension. In 1994, Boyles suggested that teachers focus only on one or two skills simultaneously, such as fluency and comprehensibility. He even wrote about giving a grade for "effort to communicate." With this particular criterion for assessment, Boyles (1994) recognized in his students the fear associated with the difficulty of oral language production. McNamara (1997) includes the social nature of oral assessment that is required of all involved in communication. Davies (1997) researched a community college French class whose goal was to teach communication skills to adults, with no college credit attached, in a non-threatening environment. Students were given a communicative task and time to prepare. They indicated to the instructor when they were ready to present. The assessment was optional. Subjects reported that the atmosphere had remained convivial, although instructors questioned the value of oral assessments in a no-credit course. Greenhough (1994) researched an oral assessment system for students of Japanese in Welsh schools. Oral assessment was flexibly timed, so assessments were gathered when students and teachers were ready. Assessments were periodic and criterion referenced, so the environment remained non-threatening. Davies, Irvine and Larrieu (1997) researched the effect of voluntary role play assessments, such as spending time with family and friends, planning a weekend holiday, or interacting with the teacher. They found that students had positive attitudes about this type of oral assessment.

Adams, Griffin and Martin (1987) proposed a partial credit model to avoid measurement problems in English as a Second Language. Students would receive partial credit for demonstrating latent language learning traits, as they approached a certain skill. Hahn, Stassen and Reschke (1989) and Moeller, Reschke and Hahn (1993) compared oral assessment outcomes between novice students of German who would be graded and students who were not. The research conclusions were that the grading of the communicative activities presented did not affect groups' overall oral proficiency. There were only individual differences. In 2001, Iwashita, McNamara and Elder tested 193 pre-university ESL learners' ability to perform narrative tasks and picture descriptions. They found that the actual task had no impact on the learner's demonstration of oral proficiency and that the task performance conditions did not influence task performance.

Regardless of the scale and means of assessment, assessment must measure the ability to use language effectively enough to attain communicative goals (Savignon, 1997). Sadly, not all teachers believe that the skill of speaking is important enough to justify the time and effort of oral assessment, causing a conflict between theory and practice (Chambers & Richards, 1995; Im, 2000).

Computer-assisted language learning can facilitate effective oral assessment. Bonk and Ockey (2003) studied the reliability and effectiveness of Rasch analysis software on peer group discussion tasks. The software studied could collect voice data for more than 1,000 subjects simultaneously. The researchers found that the oral proficiency among students increased as a result of the peer group oral assessment, and that experienced raters were consistent in their scoring of individual subjects' oral production. The rating scale included values for

54

pronunciation and overall communication. The only exception to rater reliability was in scores as they approached the ends of scale ranges. Nutta et al. (2002) found that, when textbook instruction was enhanced by technology, retention of language was greater and the students were able to produce larger chunks of language than their text-only classmates. Halleck (2007) proposed that computer simulation role play was a viable means by which to elicit a speech sample, and that simulation-based teaching may have future implications in foreign language education. Early and Swanson (2008) studied the effect of student-created digital portfolios. The intention was that the portfolios would show student progress in multiple skills including speaking. They found that the use of technology lowered the students' affective filters and that the portfolio demonstrated student progress. Additionally, as students created their digital portfolio entries simultaneously in a computer laboratory setting, individual instructional time increased.

Hallam (2008) reviewed the use of computer language laboratories. Whereas various forms of language laboratories have been in use since the mid-20th century for listening and speaking, laboratories have become multimedia settings that offer text, audio, graphics and video. Online work can be performed from any computer making distance learning possible. Hallam claimed that the language laboratory setting provided a safe and private speaking zone that would allow language learners to perfect their pronunciation, and urged that the only way to learn and retain a language was to speak it. Students, Hallam found, became immersed in language and could work at their own pace. It was suggested that it is a tool that could enhance students' foreign language acquisition and that it could have a positive impact on instruction if instructors used it thoughtfully.

Blake, Wilson, Celto & Pardo-Ballester (2008) compared traditional foreign language class oral assessment outcomes to distance-learning oral assessment outcomes in response to teacher apprehensions about the use of technology, especially when testing the oral skills. They compared "Spanish Without Walls," a Spanish I distance learning course, and a hybrid Spanish 1 course that incorporated face-to-face instructor time and internet time. The technological tools were CDs, DVDs, web pages, and synchronous bimodal chat with sound and text. All subjects were assessed orally at the end of the respective course using the same 20-minute Spanish oral test by telephone, and the researchers found that the level of oral proficiency for each group of students was comparable to first year achievement.

At the university level in English as a second language-learning, Chiu, Liou, and Yeh (2007) found that technology enhancement was particularly helpful for non-English majors. They researched the oral proficiency gains of 29 English majors and 20 non-majors who were using "CandleTalk," a web-based discourse-completion test that utilized automatic speech recognition technology.

Garcia and Laborda (2004) investigated a computer based tool, Herrameinta Informatica de Evaluacion Oral (HIEO), which can be utilized to gather voice data from up to 1,000 students simultaneously. It was a video-based tool funded by a grant from the University of Valencia, Spain, that could be applied at other universities in Europe. Lee (2007) examined the Multimedia Assisted Test of English-Speaking which was intended to establish proficiency levels among students at the Sookmyung Women's University. The assessment scale was based on the ACTFL proficiency guidelines. The Center for Applied Linguistics (CAL) was involved with the computerized assessment. The test took subjects from simple to increasingly difficult tasks and

simulated the ACTFL oral proficiency interview. Lee (2007) found that the assessment was authentic, interactive, and practical but noted that it would require further testing to measure its reliability and validity.

One K-12 public school system invested in a system of oral proficiency assessment. The Pittsburgh Public School System created a large-scale oral assessment that was based on ACTFL's Oral Proficiency Interview and the Center for Applied Linguistics' Simulated Oral Proficiency Interview. The test included a warm-up, level checks to establish starting levels, probes up to a higher level, and cool-down tasks. All students in a class could complete the 10 required items within a regular 40-minute class period. The test was successfully administered to students of multiple foreign languages, at grades 5, 8, level 3 and grade 12, over a four-year period. Fall, Adair-Hauck and Glisan (2007) reported that it was a valid, reliable, and feasible instrument for wide-scale testing. They concluded that oral proficiency would improve provided the district adopted a focus on oral proficiency as the major goal for students and teachers. Teachers were involved in the writing of open-ended questions that made up part of the assessment. Teachers were also the assessors, rating students whom they did not teach, to avoid bias. The researchers concluded that teacher involvement created buy-in. They also noted that it takes time to change perceptions and beliefs about instruction.

Language is complex. Foreign language acquisition is complex. There is no absolute in foreign language assessment, but there are standards to which assessment should be aligned. Faculties at colleges of education and faculties of foreign languages should collaborate to design and improve foreign language teacher programs. Where ACTFL and the National Council for Accreditation of Teacher Education (NCATE) program standards are both being met, there is a

close collaboration (Reese, 2008.) Universities should establish a culture of oral proficiency assessment as well as educate prospective teachers on the standards (McAlpine & Dhonau, 2007). Clark (1981) appealed to teachers to teach real communication to students, not just skills and language practice exercises. Though many classroom resources have included such skill-based "speaking practice," such textbook exercises have had more to do with pronunciation, grammar practice, and vocabulary memorization.

Teacher language qualification has been an important component in the acquisition of second language proficiency of students. The South African Norms and Standards for Teacher Education was created in 1995 by the Committee on Teacher Education Policy. The standards include outcome-based teacher requirements for linguistic competency (Nieman, 1997.) Similar proof of competence has been in place in the United States. Requirements vary by state, but the state of Florida certification includes proficiency in all five foreign language skills including oral proficiency. However, oral proficiency does not necessarily mean fluency.

In his research on the use of -target language in the classroom, Thompson (2006) compared native-speaking teachers to non-native-speaking teachers in terms of how much target language was being used in the classroom. He found that there was no statistically significant correlation between teachers' use of the target language and their native language. Further, he found that it did not have an effect on the overall language use (oral proficiency) of students. Individual differences between students, such as motivation and extraversion, would require more research.

In one study, Manley (1995) described one school district's response to the alignment of goals and assessments by creating a district-wide oral assessment system. Though lengthy, the

process was thorough and effective. First, teachers were familiarized with the need for and requirements associated with oral assessments. Teacher test-writers were selected, and they developed oral tests. The oral assessments that had been created were tested in select classrooms before widespread use was instituted. Teachers were involved throughout the process, encouraging the buy-in that is required with a change of instruction, especially a change in beliefs about instructional practices. Teachers learned first-hand what was involved in creating valid oral testing instruments and providing insight into potential difficulties with the application of oral assessments.

It is interesting to note that teachers can control their own instructional practices so they can affect, if not control, the actual learning outcomes of all of their students. Govoni (1999) studied secondary school foreign language teachers who were trained by ACTFL in the oral proficiency interview and gained their OPI certifications. The outcomes of this study were that the teachers reported being more aware of a proficiency-based curriculum. Govoni reported that the OPI certification of their teachers effected no overall change, however, in the oral proficiency of the students. Clearly, language learning involves the learner, as well as the instructor, so a variety of individual outcomes can be expected.

Purshcel (1981) observed English as a Second Language (ESL) classrooms in German schools for the conditions under which students were being trained to become orally proficient in English. He made several conclusions that remained observable in many classrooms at the time of the present study. He noted that audio equipment was used too seldom to encourage oral performance of students and that oral proficiency involved more than class participation, vocabulary use, and pronunciation (Clark, 1981). He stated that most grades were derived from

written work. Lastly, he stated that communicative competence was believed to be affected by teacher training and experience, when in fact, teacher training and experience had an insignificant influence on student oral proficiency.

In an effort to help students achieve the many stated goals of foreign language proficiency, foreign language pedagogy has been proficiency-based since the 1990s. Theoretically, it would be sound to conduct oral assessments as part of instruction for proficiency. Questions remain as to how many teachers actually apply oral assessment to their instructional methods and what causes teachers to apply oral assessment to oral proficiency instruction today.

CHAPTER 3
METHODS AND PROCEDURES

Participants

The participants in this study were Florida public middle and high school foreign

language teachers. All Florida public school teachers must hold current certification to teach a

particular foreign language, or be involved in the two-year certification process. Many private

schools have not required Florida certification, so those teachers were not included in the study

as participants. Sign language teachers were also excluded from the population. Although

considered by the Florida Foreign Language Association to be foreign language teachers, sign

language involves visual and digital skill, rather than oral and aural skills. Teachers of English as

a Foreign Language were excluded from the population. Although these teachers teach a foreign

language, the classroom experience is different because of the language immersion capability

beyond the classroom.

A list of foreign language teacher emails was compiled and alphabetized. A random

sample was selected with no stratification by district. The researcher used randomizer.org to

select random positions from a list of email addresses. Those selected email addresses became

the random sample of the population of foreign language teachers in the State of Florida. The

researcher then initiated the five research contacts that were suggested by Dillman (2000).

The method did not occur as originally proposed, and the researcher had to make several

changes during this research study. The intention had been to make contact with all foreign

language district specialists or supervisors, soliciting an email list of all middle and high school

foreign language teachers, in each district. The list would then be combined, names alphabetized,

thus creating the population from which to randomly select a sample. There were only five responses from district specialists: one response from the researcher's district, one response with a link to the district website from which to create an email list, manually, one from a district stating that the district did not distribute email addresses and two from districts with a research approval process. The paperwork was completed for those latter two districts. There was no response from one district and the other declined to allow the study. One district foreign language contact advised that she would invite her teachers to email the researcher if they were interested in participating. She could not be made to understand that this would not be a random sample. Being in one of the largest districts in the state, the researcher used the website to access all schools and almost all foreign language teacher email addresses. Receiving no response from the remaining district specialists, the sample had to be generated, manually.

Using the Florida Department of Education (DOE) enrollment, by district (Appendix B), the researcher sorted the list according to student enrollment, largest to smallest. There were 67 districts, excluding those Department of Education (DOE) "districts" that were limited to affiliation with a state university.

The researcher accessed districts' web sites, isolate middle and high school web sites, locate their foreign language teachers and create an email distribution list. The researcher's district was the 17th largest in the state, and the researcher included it, due to ease of access. The researcher compiled the names from one of the largest districts then moved to the smallest districts. Several of the lesser-populated districts did not have any foreign language teachers, so not all districts became part of the population to be studied.

The most recent student enrollment data available at the time of the study was for 2007-2008 and reflected a student population for all Florida public schools of 2,646,115 distributed among 67 districts. It was not reasonable, possible, or necessary to retrieve all foreign language email addresses from the 67 districts, so the researcher selected one district from at most every four based on increasing student membership and geographic location.

One school district that was initially selected used a password-protected website exclusively, so it did not offer public access to teacher email addresses. The researcher passed to a similarly-sized neighbor. Similarly, if a district did not have any foreign language teachers, the researcher passed to neighboring districts with increasing student enrollments until a foreign language teacher was identified.

Of the total state student membership of 2,646,115, the districts included in this study had a membership of 672,739, which was 25% of the total Florida student membership. The researcher included 17 of the 67 districts in the population, 25% of the total number of school districts. The researcher was satisfied that this was an adequate representation of the population studied.

Based on the number of foreign language teachers in one quarter of the districts, the population was estimated at 478 x 4, which was 1,912 middle and high school foreign language teachers. Although several schools did not allow public access to teachers' email addresses or subject area taught, the researcher was confident that 1,912 was a reasonable estimate of the population. The maximum score on the questionnaire was 83, the minimum was 0. Thus, $(83 - 0) / 4 = 20.75$ was the estimate of the population standard deviation. Squared, 430.56 was

the population variance according to Tchebbysheff's Theorem. The researcher applied a margin of error (B) of 3.0

$$D = B^2/4 = (3.0)^2/4 = 2.25$$

$$n = \frac{N\sigma^2}{(N-1)D+\sigma^2} = \frac{1,912\ (20.75)^2}{(1,911)2.25+(20.75)^2} = \frac{823,235.5}{4730.3} = 174.03$$

The districts (and their student memberships) included in the study were: Charlotte (17,779), Marion (42,565), Sumter (7,518), Volusia (64,570), Flagler (12,774), Walton (6,967), St. Johns (27,867), Broward (258,895), Liberty (1,513), Pinellas (107,895), St. Lucie (40,347), Hardee (5,014), Gilchrist (2,889), Gulf (2,171), Columbia (10,134), Pasco (66,313) and Osceola (52,742). Holmes (3,430), Taylor (3,389) and Jefferson (1,154) had no foreign language teachers identified in their middle and high schools so they were not included. Nassau County (11,079) used a password-protected website, so the researcher was not able to access those teachers' email addresses. The researcher included Pasco County (66,313), instead, despite the larger student enrollment, because there appeared to be a lack of inclusion from that area of the state. In addition to student enrollment, the researcher was mindful of geographical variety in the selection of school districts so as to ensure geographic diversity across the state of Florida.

In several districts, some schools did not allow public access to teacher emails because they used a parent portal. Several schools did not include subject area identification for their teachers, so the researcher was not able to isolate their foreign language teachers, if there were any, at those mostly middle schools. There was one school with an inactive website, and teachers' email addresses could not be accessed. These individual school exceptions from the population (and the appropriate district) were: Hollywood (Broward), MacArthur (Broward),

Gibbs High School (Pinellas), Lakewood (Pinellas), Northeast High School (Pinellas), Pinellas Park (Pinellas), Meadowlawn (Pinellas), Morgan Fitzgerald (Pinellas), Oak Grove (Pinellas), Tarpon Springs Middle School (Pinellas), Ft. Pierce Westwood (St. Lucie), Lincoln Park Academy (St. Lucie), Port St. Lucie High School (St. Lucie), Southern Oaks Middle School (St. Lucie), Columbia High School (Columbia), Liberty High School (Osceola), Poinciana High School (Osceola), Anclote High School (Pasco), Fivay (Pasco), Ridgewood (Pasco), Charles Rush Middle School (Pasco), Chasco Middle School (Pasco), Pine View Middle School (Pasco) and Seven Springs (Pasco).

Pasco County ultimately did not allow research to be conducted in the district based on improper protocol on the part of the researcher. The researcher had inadvertently made contact without awaiting district approval to conduct research. At first contact to the randomly-selected population sample, the researcher received an email from one of the district personnel, requesting that no further contact be made with those employees. Further, the district person had replied to all participants, regardless of district, that Pasco county employees must not participate in the study because permission had not been granted. The researcher was concerned that this alert would negatively impact the others' willingness to participate. The random sample had included 21 teachers from that district. They were removed from the list and replaced with another 21 random participants, generated using randomizer.org, from the other districts. Although 17 districts were selected out of the 67, research was conducted in 16 of them.

Three individual teachers replied on the day of first contact that they would not be participating in the research. Another three random emails were selected. Any time a name was

replaced, whether by the non-participating district or the three individuals who declined to participate, randomizer.org was used to randomly select numbers from the remaining emails.

At second contact, the researcher received another three email responses, letting her know that they would not be participating. They were replaced, and first contact was made with the replacements. A total of 27 participants were replaced, altogether, with three additional random selections from the remaining population.

It was a challenging task to keep track of all of the participants, moving email addresses from one list to another and making the appropriate next contact, since four different groups of participants emerged. The first group was the original population sample;, the second group was the Pasco county replacements;, the third group was comprised of the first contact refusals; the fourth group consisted of the second contact refusals. Mindful that the end of the K-12 school year was quickly approaching, the researcher made the next contact whenever a weekend passed with no more responses from that group. She kept spreadsheets and notes, replying as soon as possible to her respondent-colleagues.

The researcher received helpful personal feedback from her colleagues. Though several teachers were unwilling, for whatever reasons, to participate in the study, they were courteous enough to respond to either the first or second email contact. This allowed the researcher to randomly select replacements from the remaining population and still complete the study. Almost half of the teachers included some sort of personal comments and well-wishes at the end of the questionnaire.

<center>Materials</center>

The researcher created an online questionnaire (Appendix C), according to the standards of Dillman (2000). On this questionnaire, teachers were asked to indicate the amount of instructional time dedicated to reading, writing, listening, speaking the foreign language, and learning about the cultures associated with the foreign language being taught. Respondents were asked to indicate the distribution of those same skills on assessments. Respondents were asked to identify resources that were available to them, their self-reported level of fluency in the language taught, their level of education, teaching experience and school demographics. Many of the questionnaire items were written according to cognitive design to improve respondent recall. Classroom applications of specific foreign language practices were listed to increase respondent recall and response accuracy.

All respondents needed to have an email address and access to the internet in order to complete the questionnaire, and all public school employees have had access to both. The researcher sent the questionnaire to teachers' professional email accounts, so that they were able to access the embedded link to the internet questionnaire from school. At first contact, one teacher provided her personal email and indicated that her preference was to be contacted at home, which the researcher did. All others responded from their professional email accounts.

<center>Procedures</center>

The researcher conducted this quantitative research study which was initiated only after the research was approved by the Institutional Review Board of the University of Central Florida (Appendix D). The dependent variable was the application of oral assessment in Florida public

<center>67</center>

foreign language classrooms. The independent variables included the available teacher resources, level of fluency of teachers, their level of education, teaching experience, and school demographics. All variables were measured using data obtained on the questionnaire. Respondents identified their instruction and assessment of the foreign language skills, according to a series of six-item Likert-type scale questions. The remaining items were answered using a Yes or No response.

Applying Dillman's (2000) principles of survey research, initial contact was a pre-notice email to respondents advising them of an upcoming survey. Second contact was an email with the questionnaire embedded so that respondents would click on the email and be directed immediately to the questionnaire. Respondents were asked to send a separate email, letting the researcher know that they have completed the survey, so that their names could be removed from the follow-up list. Respondents were able to navigate within the survey. Series of questions with similar response scales were grouped, demographic questions were placed at the end of the instrument, and all questions were numbered.

At the first contact, the researcher let participants know that they had been randomly selected and would be receiving a link to an online foreign language instruction questionnaire. Several people responded enthusiastically. Second contact included the questionnaire link, embedded in the email. The researcher was inundated with email responses regarding completion on that first day of the questionnaire administration. Several people included recommendations for future research. Many included well-wishes for a successful study and school year. Additional responses came in for three days, but no responses followed. As subjects responded

via email that they had completed the questionnaire, their names were removed from the email follow-up distribution list.

The researcher initiated contact number three to the remaining names on the distribution list. Several participants emailed that they had already completed the questionnaire but had neglected to email until third contact. Each of the next three days, emails were received from participants. The researcher continued with contact number four. At each contact, the link was embedded, to encourage response from the participants.

Contact number five was made with the remaining teachers on the last day of school. The researcher added a personal note about going to Ohio to grade this year's AP French tests and that she hoped to start compiling results upon her return in eight days. She hoped that this personal detail would encourage colleagues to help her to complete her research. Several more questionnaires were completed during the final eight-day period.

Once all data were collected, they were imported to an excel document and entered into SPSS. They were analyzed with a paired t-tests and a series of ANOVA's. One question was qualitative in nature, and participant responses were coded.

CHAPTER 4
DATA ANALYSIS

Research Question 1

What is the actual allocation of instructional time and assessment on each of ACTFL's five elements of foreign language proficiency?

On the Florida Middle and High School Foreign Language Questionnaire (Appendix A), questions 1-10 addressed the allocation of instructional time and assessment on each of the elements of foreign language proficiency. These were reading, writing, speaking, listening, and cultures of the peoples who speak the language being taught.

Two responses were given by each respondent to address the amount of time allocated to each foreign language skill. Each participant responded to a question concerning the amount of instructional time typically dedicated to each of the five skills. Additionally, each participant responded to a question concerning the amount of unit assessments typically dedicated to each of the same five skills. Therefore, each participant provided two responses to questions about each of the five skills, and the data were analyzed using a paired samples t-test.

Null hypotheses:

1. $H_0 : \mu_{InstructRead} = \mu_{AssessRead}$
   $H_1 : \mu_{InstructRead} \neq \mu_{AssessRead}$

2. $H_0 : \mu_{InstructWrite} = \mu_{AssessWrite}$
   $H_1 : \mu_{InstructWrite} \neq \mu_{AssessWrite}$

3. $H_0 : \mu_{InstructSpeak} = \mu_{AssessSpeak}$
   $H_1 : \mu_{InstructSpeak} \neq \mu_{AssessSpeak}$

4. $H_0 : \mu_{InstructListen} = \mu_{AssessListen}$
   $H_1 : \mu_{InstructListen} \neq \mu_{AssessListen}$

5. $H_0 : \mu_{InstructCulture} = \mu_{AssessCulture}$
   $H_1 : \mu_{InstructCulture} \neq \mu_{AssessCulture}$

Table 1

Instruction vs. Assessment Mean Allocations of Instructional Time

	Paired Samples	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Instr.Read	2.24	79	1.089	.122
	UnitAssessRead	2.56	79	1.268	.143
Pair 2	Instr.Write	2.32	78	1.157	.131
	UnitAssessWrite	2.31	78	1.036	.117
Pair 3	Instr.Speak	1.91	79	.963	.108
	UnitAssessSpeak	1.47	79	.845	.095
Pair 4	Instr.Listen	2.92	79	1.118	.126
	UnitAssessListen	1.15	79	1.014	.114
Pair 5	Instr.Culture	1.61	79	.838	.094
	UnitAssessCulture	1.16	79	.791	.089

Regardless of the statistical significance of these outcomes, it bears mentioning that the differences between four of these means were slight. The mean teacher responses for instruction and assessment of reading were 2.24 and 2.56, respectively. A response of 2 was "around 25%" and a response of 3 was "around 50%." The mean teacher responses for instruction and assessment of writing were 2.32 and 2.31, respectively. Although being statistically significant, allowing the rejection of the null hypothesis that the instruction and assessment are aligned, the means were very close. The mean teacher responses for instruction and assessment of speaking were 1.91 and 1.47, respectively. A response of 1 was "less than 25%" and 2 was "around 25%." Again, these means were very close and indicated that teachers were reporting that the instruction and assessment were, on average, similar. The mean teacher responses for instruction and assessment of listening were 2.92 and 1.15. The difference was statistically significant, so

the null hypothesis that the instruction and assessment of listening are equal was rejected. These means were quite different. Almost 50% of instructional time was dedicated to listening and less than 25% of unit tests were dedicated to listening. Lastly, the mean teacher responses for instruction and assessment of cultures were 1.61 and 1.16, respectively, somewhere between "less than 25%" and "around 25%."

Table 2
Instruction vs. Assessment: Five Foreign Language Skills

| | | | | Paired Differences | | | | |
| | | | | 95% Confidence Interval of the Difference | | | | |
Paired Samples	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1 Instr.Read - UnitAssessRead	-.316	1.161	.131	-.576	-.056	-2.423	78	.018
Pair 2 Instr.Write - UnitAssessWrite	.013	1.233	.140	-.265	.291	.092	77	.927
Pair 3 Instr.Speak - UnitAssessSpeak	.443	.944	.106	.232	.654	4.172	78	.000
Pair 4 Instr.Listen - UnitAssessListen	1.772	1.510	.170	1.434	2.110	10.430	78	.000
Pair 5 Instr.Culture - UnitAssessCulture	.443	.902	.102	.241	.645	4.365	78	.000

$p < .05$

Null hypotheses:

1. $H_0 : \mu_{InstructRead} = \mu_{AssessRead}$
      $H_1 : \mu_{InstructRead} \neq \mu_{AssessRead}$

A paired samples *t* test revealed a statistically reliable difference between the mean amount of instructional time dedicated to reading (M = 2.24, s = 1.09) and the mean amount of

assessment dedicated to reading (M = 2.56, s = 1.27), $t(78) = 2.423$, $p = .018$, $\alpha = .05$. Thus, the

null hypothesis that the amount of instruction dedicated to reading is equal to the amount of

assessment dedicated to reading the target language was rejected.

2. $H_0 : \mu_{\text{InstructWrite}} = \mu_{\text{AssessWrite}}$
  $H_1 : \mu_{\text{InstructWrite}} \neq \mu_{\text{AssessWrite}}$

A paired samples t test failed to reveal a statistically reliable difference between the mean

amount of instructional time dedicated to writing (M = 2.32, s = 1.16) and the mean amount of

assessment dedicated to writing (M = 2.31, s = 1.04), $t(77) = 0.092$, $p = .927$, $\alpha = .05$. The null

hypothesis that the amount of instruction dedicated to writing is equal to the amount of

assessment dedicated to writing the target language could not be rejected.

3. $H_0 : \mu_{\text{InstructSpeak}} = \mu_{\text{AssessSpeak}}$
  $H_1 : \mu_{\text{InstructSpeak}} \neq \mu_{\text{AssessSpeak}}$

A paired samples t test revealed a statistically reliable difference between the mean

amount of instructional time dedicated to speaking (M = 1.91, s = 0.96) and the mean amount of

assessment dedicated to speaking (M = 1.47, s = 0.85), $t(78) = 4.172$, $p < .001$, $\alpha = .05$. The null

hypothesis that the amount of instruction dedicated to speaking is equal to the amount of

assessment dedicated to speaking in the target language was rejected.

4. $H_0 : \mu_{\text{InstructListen}} = \mu_{\text{AssessListen}}$
  $H_1 : \mu_{\text{InstructListen}} \neq \mu_{\text{AssessListen}}$

A paired samples t test revealed a statistically reliable difference between the mean

amount of instructional time dedicated to listening (M = 2.92, s = 1.12) and the mean amount of

assessment dedicated to listening (M = 1.15, s = 1.01), $t(78) = 10.430$, $p < .001$, $\alpha = .05$. The null

hypothesis that the amount of instruction dedicated to listening is equal to the amount of assessment dedicated to listening to the target language was rejected.

5. $H_0 : \mu_{InstructCulture} = \mu_{AssessCulture}$
$H_1 : \mu_{InstructCulture} \neq \mu_{AssessCulture}$

A paired samples t test revealed a statistically reliable difference between the mean amount of instructional time dedicated to culture (M = 1.61, s = 0.84) and the mean amount of assessment dedicated to culture (M = 1.16, s = 0.79), $t(78) = 4.365$, $p < .001$, $\alpha = .05$. Thus, the null hypothesis that the amount of instruction dedicated to culture is equal to the amount of assessment dedicated to the cultures associated with the target language was rejected.

There was insufficient evidence to conclude that the amount of instruction and assessment of writing the target language was different. There was sufficient evidence to conclude that the amounts of instruction and assessment of the other four skills of reading, speaking, listening, and culture were different.

The null hypothesis that the amount of instructional time dedicated to reading, speaking, listening and learning about cultures is equal to the amount of unit tests dedicated to the same foreign language skills was rejected. This means that the stated goals for learning those skills were not in alignment with the assessments of the same individual skills.

The null hypothesis that the amount of instructional time dedicated to writing is equal to the amount of unit tests dedicated to the skill of writing was not rejected. This means that the stated goal for writing was in alignment with the assessment of writing.

Teacher-respondents were asked to self-report on the amount of typical unit tests dedicated to each of the five foreign language skills. In a comparison of these five means,

reading was the most-assessed skill. It consumed, on average, 39% of unit tests. Writing

followed reading at about 32% of unit tests. Speaking, listening, and culture were reported as

consuming less than 25% of unit tests. Based on the amount of assessment dedicated to reading

and writing, it was assumed that those were the most valued skills. However, it is worth noting

that speaking was assessed, on average, more than listening and culture.

To examine the frequency of the amount of each unit test dedicated to each of the five

foreign language skills, each of the five skills were treated individually. The frequency data for

each of the five skills are displayed in histograms.

First, the frequency with which reading assessment is administered on unit tests was

examined. Teachers self-reported the amount of unit tests typically dedicated to reading target

language. This information is displayed in Table 3.

Table 3
Frequency of Varying Amounts of Reading Assessment Administered on Unit Tests

	UnitAssessRead	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	2	1.5	2.5	2.5
	Less than 25%	11	8.3	13.9	16.5
	Around 25%	33	24.8	41.8	58.2
	Around 50%	17	12.8	21.5	79.7
	Around 75%	6	4.5	7.6	87.3
	More than 75%	10	7.5	12.7	100.0
	Total	79	59.4	100.0	
Missing	System	54	40.6		
Total		133	100.0		

Table 3 illustrates that 33 (42%) of the foreign language teachers surveyed dedicate around 25% of unit tests to reading assessment. The mean of 2.56 indicated that the average amount of a unit test dedicated to reading assessment was between 25% and 50% of the test.

A histogram (Figure 1) shows the frequency of varying amounts of unit tests dedicated to reading assessment.
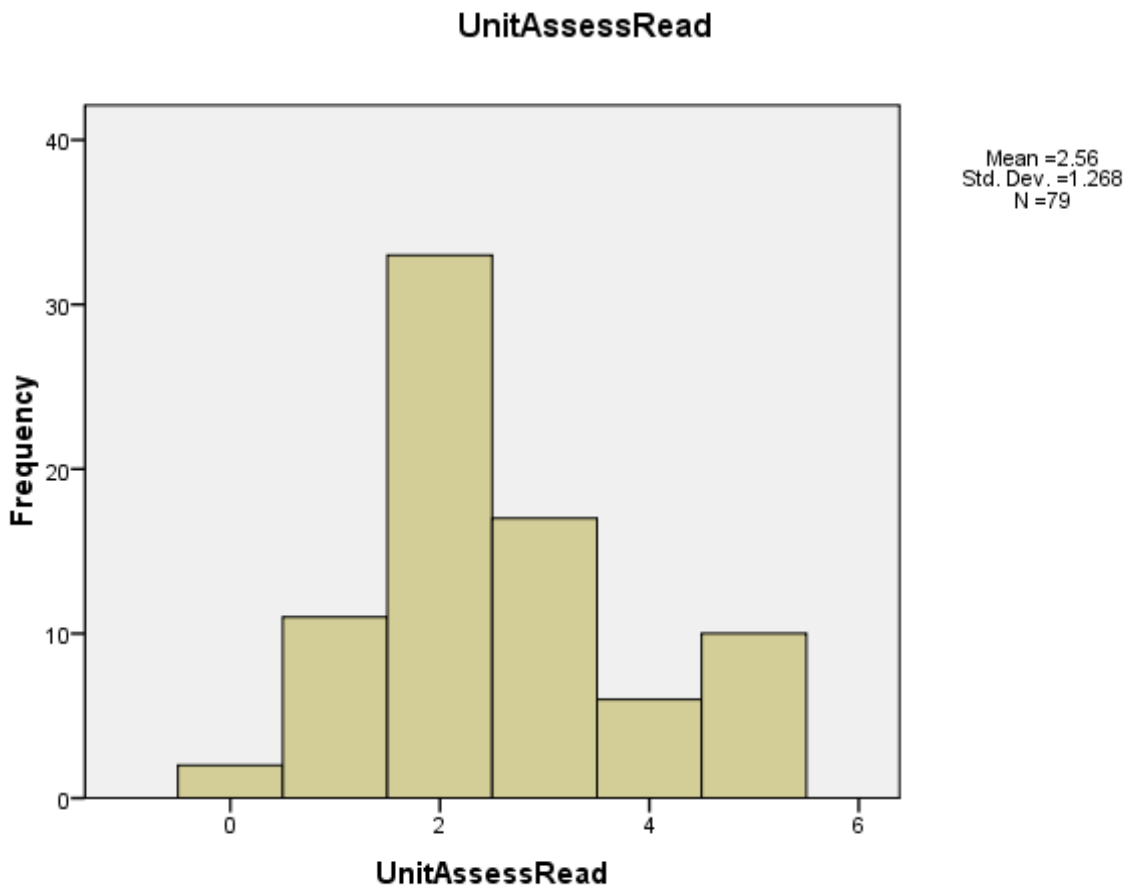

Figure 1. Frequency of varying amounts of reading assessment administered on unit tests.

Figure 1 shows that most teachers reported reading assessment as consuming about 25% of typical unit tests. A smaller group of teachers reported that reading assessment consumed more than 25% of typical unit tests. A similar number of teachers reported that reading assessment consumed more than 75% of unit tests or less than 25% of unit tests. Few teachers reported dedicating around 75% of typical unit tests to reading assessment, and even fewer reported reading assessment as less than 25% of unit tests.

Table 4 presents the frequency with which student writing was assessed on unit tests. Teachers self-reported the amount of unit tests typically dedicated to students' writing in target language.

Table 4

Frequency of Varying Amounts of Written Assessment Administered on Unit Tests

	UnitAssessWrite	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than 25%	17	12.8	21.5	21.5
	Around 25%	36	27.1	45.6	67.1
	Around 50%	15	11.3	19.0	86.1
	Around 75%	8	6.0	10.1	96.2
	More than 75%	3	2.3	3.8	100.0
	Total	79	59.4	100.0	
Missing	System	54	40.6		
Total		133	100.0		

Table 4 illustrates that 36 (45%) of the foreign language teachers surveyed dedicated around 25% of unit tests to assessment of writing. The mean of 2.29 indicates that the average amount of a unit test dedicated to written assessment was between 25% and 50%. This was lower than the mean amount of assessment dedicated to reading. A histogram (Figure 2) shows the

frequency of varying amounts of unit tests dedicated to the assessment of writing in the target

language.
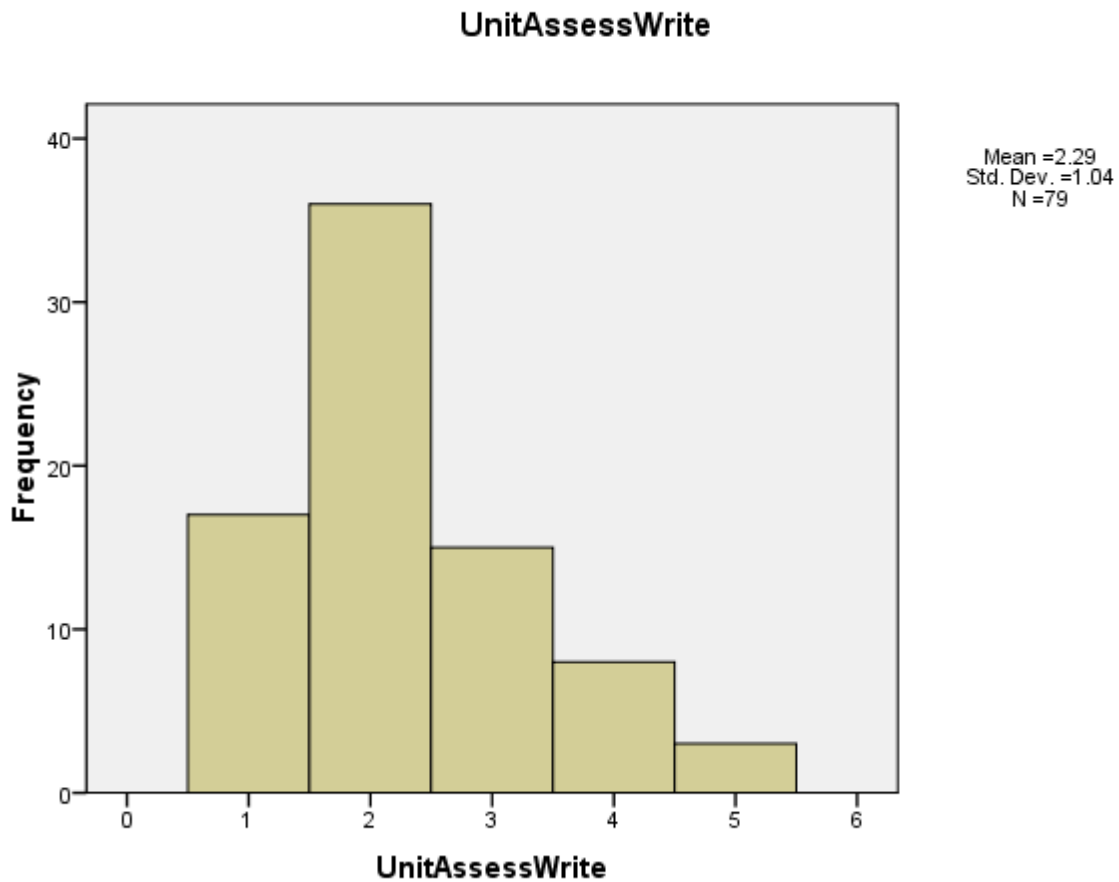
Figure 2. Frequency of varying amounts of written assessment administered on unit tests.

Figure 2 shows that most teachers reported written assessment consuming about 25% of

typical unit tests. A smaller group of teachers reported that written assessment consumed less

than 25% of typical unit tests. Decreasing numbers of teachers reported dedicating around 50%,

around 75% and more than 75% of unit tests to written assessment. It would appear that more

teachers dedicated a larger portion of their unit tests to the assessment of students' reading skills than their writing skills.

Table 5 presents data on the frequency with which students are assessed on speaking target language on typical unit tests. Teachers self-reported the amount of unit tests typically dedicated to students' oral production in target language.

Table 5
Frequency of Varying Amounts of Oral Assessment Administered on Unit Tests

	UnitAssessSpeak	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	7	5.3	8.9	8.9
	Less than 25%	35	26.3	44.3	53.2
	Around 25%	33	24.8	41.8	94.9
	Around 50%	2	1.5	2.5	97.5
	Around 75%	1	.8	1.3	98.7
	More than 75%	1	.8	1.3	100.0
	Total	79	59.4	100.0	
Missing	System	54	40.6		
Total		133	100.0		

Table 5 illustrates that 35 (44%) of the foreign language teachers surveyed dedicate less than 25% of unit tests to assessment of speaking. Closely following, 33 (42%) of the teachers dedicate around 25% of unit tests to the assessment of speaking. The mean of 1.47 indicated that the average amount of a unit test dedicated to oral assessment was less than reading or writing, somewhere between "less than 25%" and "around 25%." Seven teachers (9%) reported not administering any oral assessment on unit tests. One teacher (1%) reported oral assessment at 50% of unit tests, and one teacher reported oral assessment at 75% of unit tests. A histogram (Figure 3) shows the frequency of varying amounts of unit tests dedicated to oral assessment.

79

UnitAssessSpeak

Mean =1.47
Std. Dev. =0.845
N =79

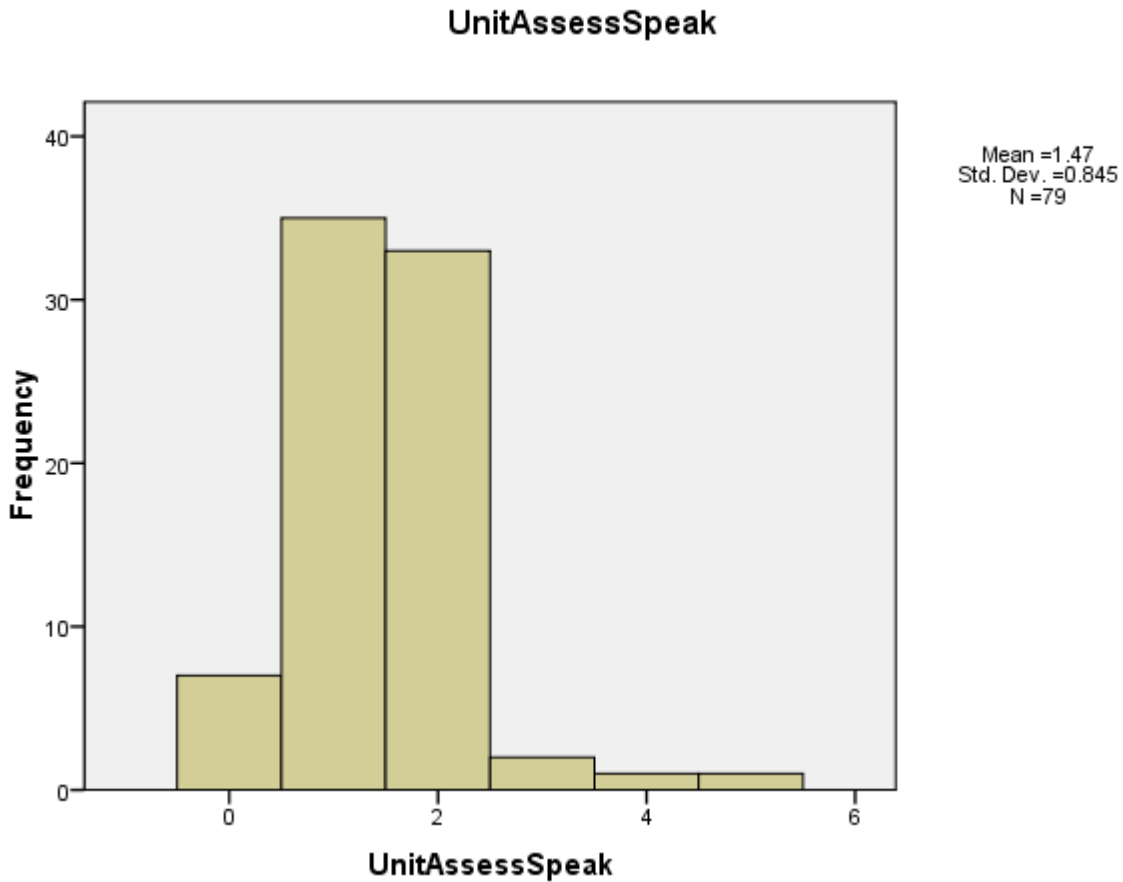*Figure 3*. Frequency of varying amounts of oral assessment administered on unit tests.

Figure 3 indicates that almost all teachers reported oral assessment as consuming either less than or around 25% of typical unit tests. A noticeably smaller group reported not conducting any oral assessment on unit tests. Even fewer reported oral assessment at around 50%, around 75%, and more than 75% of unit tests.

Table 6 contains the analysis of the frequency with which students are assessed on their listening comprehension on typical unit tests. Teachers self-reported the amount of unit tests typically dedicated to students' listening comprehension of target language.

Table 6

Frequency of Varying Amounts of Listening Assessment Administered on Unit Tests

	UnitAssessListen	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	23	17.3	29.1	29.1
	Less than 25%	30	22.6	38.0	67.1
	Around 25%	20	15.0	25.3	92.4
	Around 50%	3	2.3	3.8	96.2
	Around 75%	3	2.3	3.8	100.0
	Total	79	59.4	100.0	
Missing	System	54	40.6		
Total		133	100.0		

Table 6 illustrates that 30 (38%) of the foreign language teachers surveyed dedicated around 25% of unit tests to assessment of listening comprehension. The mean of 1.15 indicates that the average amount of a unit test dedicated to the assessment of listening comprehension was less than 25%. This was the lowest of the assessment means. Perhaps this assessment was taking place elsewhere than on a unit test. It was interesting to note that the mean amount of unit tests dedicated to oral assessment was greater than the mean amount of unit tests dedicated to listening assessment. A histogram (Figure 4) shows the frequency of varying amounts of unit tests dedicated to listening comprehension of target language.

UnitAssessListen
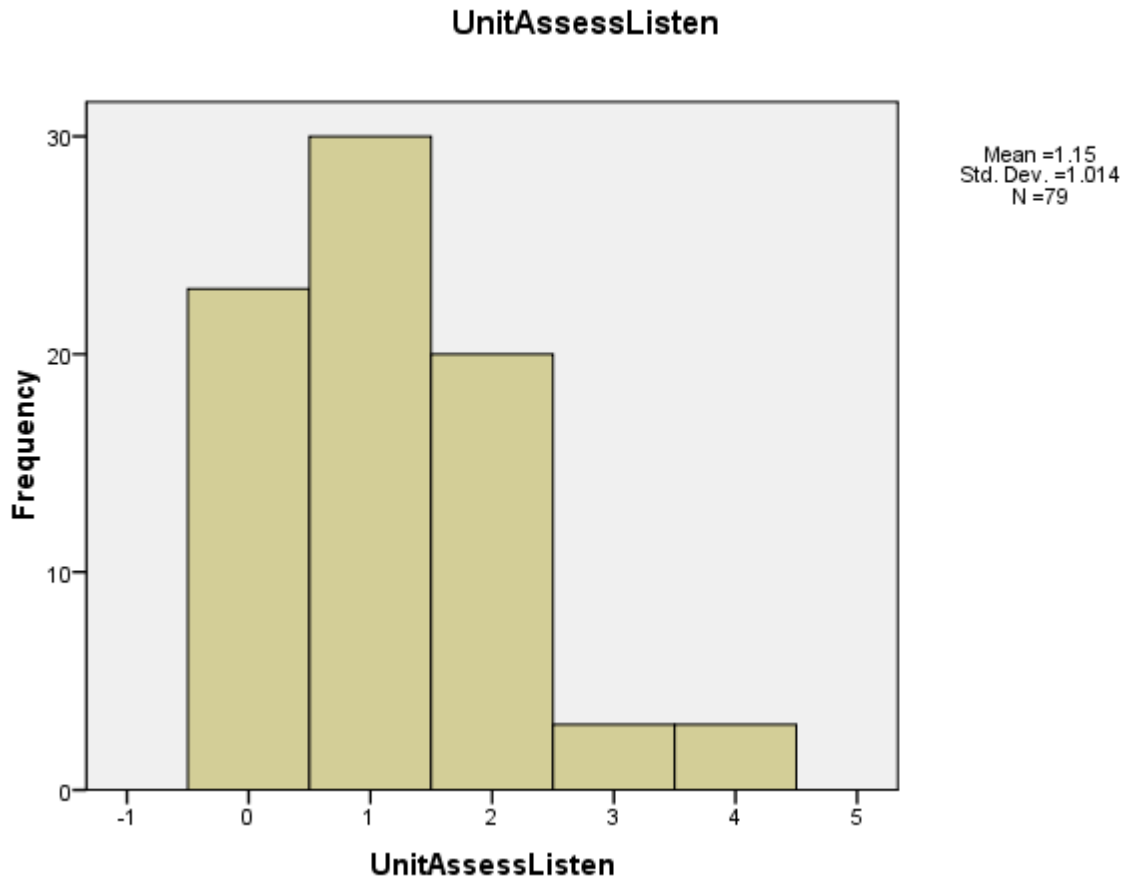
Figure 4. Frequency of varying amounts of listening assessment administered on unit tests.

Figure 4 shows that most teachers reported that listening comprehension takes up less than 25% of typical unit tests. The second largest group reported not including listening comprehension on unit tests. Perhaps students' listening comprehension has been measured by teachers in another type of assessment. The third largest group of teachers reported listening comprehension at about 25% of unit tests. The two smallest groups reported listening comprehension at around 75%, and more than 75% of unit tests.

Table 7 presents the frequency with which students are assessed on their knowledge of the cultures that are associated with the speakers of the target language. Teachers self-reported the amount of unit tests typically dedicated to students' cultural knowledge.

Table 7

Frequency of Varying Amounts of Cultural Assessment Administered on Unit Tests

	UnitAssessCulture	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	11	8.3	13.9	13.9
	Less than 25%	49	36.8	62.0	75.9
	Around 25%	16	12.0	20.3	96.2
	Around 50%	2	1.5	2.5	98.7
	More than 75%	1	.8	1.3	100.0
	Total	79	59.4	100.0	
Missing	System	54	40.6		
Total		133	100.0		

Table 7 indicates that 49 (62%) of the foreign language teachers surveyed dedicated less than 25% of unit tests to assessment of cultural knowledge. The mean of 1.16 indicates that the average amount of a unit test dedicated to cultural knowledge assessment was less than 25%. This mean was very close to that of the assessment of listening comprehension. Compared to reading, writing and speaking assessments, according to the responses given, listening comprehension and cultural knowledge assessments consumed a lower portion of unit tests A histogram (Figure 5) shows the frequency of varying amounts of unit tests dedicated to cultural knowledge.

UnitAssessCulture

Mean =1.16
Std. Dev. =0.791
N =79
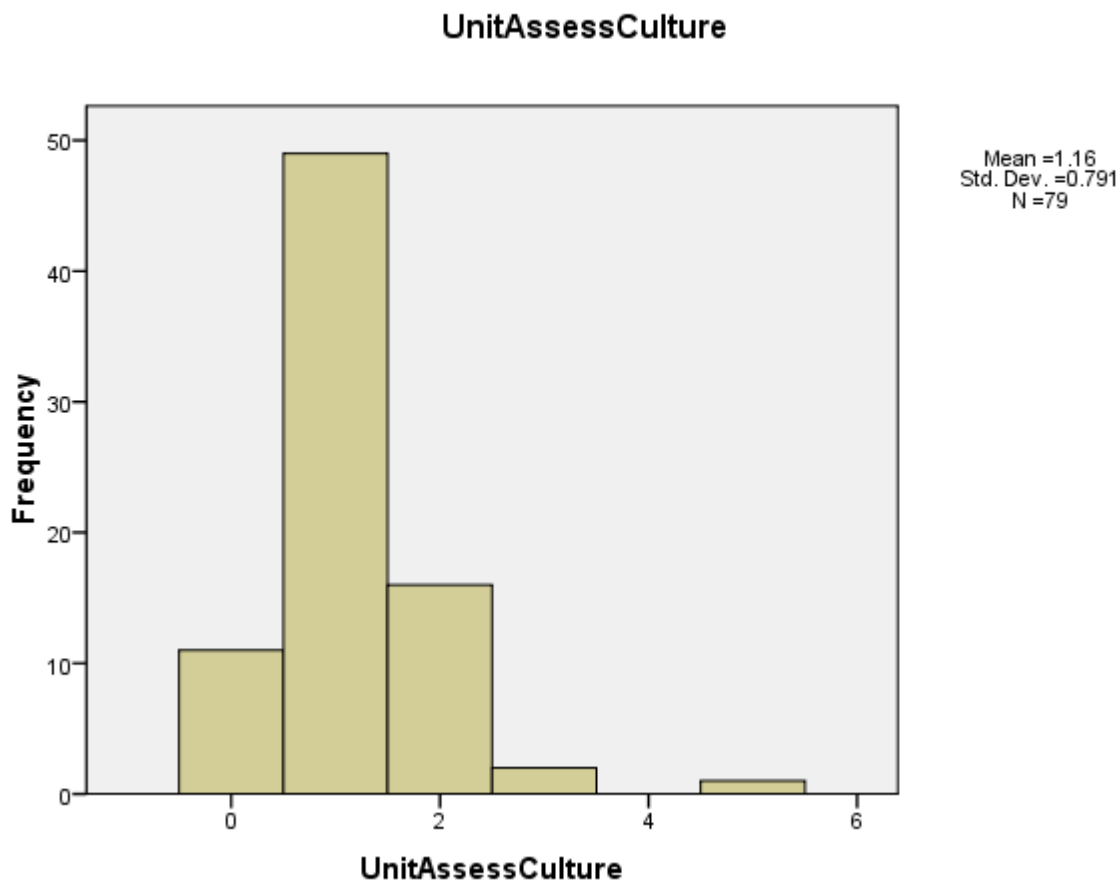
Figure 5. Frequency of varying amounts of cultural assessment on unit tests.

A histogram shows that most teachers reported that a demonstration of cultural knowledge consumes less than 25% of typical unit tests. A much smaller group reported assessing the cultural aspect of foreign language study at around 25%. A slightly smaller group reported no assessment of culture on unit tests. A few teachers reported the assessment of cultural knowledge at around 50% and more than 75% of unit tests.

Questions11-25 asked respondents whether or not they perform various, other assessments, as part of their foreign language instruction. These assessments might be considered to be less formative than the previous questions, which related to typical unit tests. Respondents answered "yes" or "no" to a series of questions concerning their application of these assessments of the five foreign language skills of reading, writing, speaking, listening, and cultural knowledge.

In declining order of most commonly performed, they are: reading comprehension (100%), listening comprehension with teacher's voice (97%), written fill-in-the-blanks (96%), written paragraphs (91%), speaking presentations (89%), speaking dialogues (89%), listening comprehension with CD (87%), speaking interviews (82%), speaking responses to written prompts/questions (77%), written definitions of vocabulary (72%), speaking responses to picture prompts (70%), speaking and being graded with an observational checklist (44%), speaking student self-assessment (35%), speaking and writing portfolio of work samples (28%), and speaking with computer-assisted voice data collection (19%). This indicated that these assessments were being applied in foreign language classrooms. They did not necessarily appear on unit tests.

<u>Research Question 2</u>

Does the allocation of textbook and computer resources have an impact on the application of oral proficiency assessment?

On the Florida Middle and High School Foreign Language Questionnaire (Appendix A), questions 26-28 addressed the resources that are available to each respondent. These include a

teacher resource textbook package, regular access to a computer laboratory, and the availability of headsets with microphones in the computer laboratory. Regular access was defined on the questionnaire as at least once per week.

The availability of teacher resources and computer laboratory were two independent variables. All 79 respondents stated that they had teacher resource packages, so this was not a variable. The researcher did not include in the results question 28, asking if the computer laboratory had headsets with microphones. Of the respondents, 26 stated that they had access to a computer laboratory, but 30 respondents stated that those computer laboratories had headsets with microphones. There were 4 more respondents with computer laboratories with headsets than respondents with computer laboratories, so the question must have been misunderstood. There was one dependent variable, the amount of oral assessment on unit tests. It is interesting to note that only 26 respondents (33%) reported having regular weekly access to a computer laboratory. A total of 53 respondents (67%) reported not having regular weekly access to a computer laboratory. The researcher performed a univariate ANOVA. The results are displayed in Table 8.

Table 8

Between-Subjects Effects: Availability of a Computer Laboratory on Oral Assessment

| | Dependent Variable: UnitAssessSpeak | | | | |
Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	.457[a]	1	.457	.637	.427
Intercept	156.153	1	156.153	217.767	.000
Computer Lab	.457	1	.457	.637	.427
Error	55.214	77	.717		
Total	226.000	79			
Corrected Total	55.671	78			

[a]R Squared = .008 (Adjusted R Squared = -.005)

p < .05

These results indicated that the variable, Computer Lab, was not statistically significant $F(1,78) = .637$, $p = 0.427$. There was no difference in the means among the two groups (teachers with access to a computer laboratory once per week and teachers without such regular access). Whether or not teachers had access to a laboratory did not have an effect on the amount of oral assessment that they conducted on a unit test.

Research Question 3

Does the level of experience of the teachers, identified as (a) number of years teaching foreign language and (b) level of education, have an impact on the application of oral assessment?

On the Florida Middle and High School Foreign Language Questionnaire (Appendix A), questions 40 and 41 addressed the level of experience of each respondent. Experience was identified as a certain range of years teaching foreign language and a certain level of education.

There were two independent variables. The years of teaching experience had five groups. The questionnaire had four groups within the independent variable level of education. There was

one dependent variable, the amount of oral assessment on unit tests. The researcher performed a univariate ANOVA.

Because the group "doctoral degree" had no respondents and the group "Ed.S." had two respondents, the researcher collapsed that group. Level of education was either "bachelor's degree" or "graduate degree." The second independent variable, years of teaching experience, had five groups. As there were only six respondents within the group "7-10 years," the researcher ran a Levene's test of equity to compare the variances among the groups to test for homogeneity of variance. The results of the Levene's test for homogeneity of variance are displayed in Table 9.

Table 9

Test for Homogeneity of Variance Among Teaching Experience Groups: UnitAssessSpeak

Levene Statistic	df1	df2	Sig.
2.470	4	74	.052

$p < .05$

A Levene's test revealed that homogeneity of variance can be assumed. Therefore, the researcher ran the univariate ANOVA with two groups for level of education and the original five groups for years of teaching experience. The results are displayed in Table 10. Of the respondents, 43 (54%) possessed a bachelor's degree, and 36 respondents (46%) possessed a graduate degree of some kind. In terms of teaching experience, 14 respondents (18%) had 1-3 years of experience, 15 respondents (19%) had 4-6 years of experience, six respondents (8%) had 7-10 years of experience, 20 respondents (25%) had 11-20 years of experience, and 24

88

respondents (30%) had more than 20 years of experience. It is interesting to note that more than

half of the respondents had 11 or more years of teaching experience.

Table 10

Between-Subjects Effects of Education and Experience on the Administration of Oral Assessment on Unit Tests

	Dependent Variable: UnitAssessSpeak				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.504[a]	9	1.056	1.578	.139
Intercept	113.690	1	113.690	169.920	.000
HighestDegree	1.308	1	1.308	1.954	.167
YearsTeaching	2.073	4	.518	.775	.545
HighestDegree * YearsTeaching	7.578	4	1.895	2.832	.031
Error	46.167	69	.669		
Total	226.000	79			
Corrected Total	55.671	78			

[a]R Squared = .171 (Adjusted R Squared = .063)
$p < .05$

The interaction between the variables Highest Degree and Years Teaching was

statistically significant: $F_{(4,78)} = 2.832$, $p = 0.031$, and the null hypothesis was rejected. There

was an interaction between the independent variables of the level of education of the teacher and

years of experience. The effects of the variables Highest Degree and Years Teaching were not

significant: $F_{(1,78)} = 1.954$, $p = .167$ and $F_{(4,78)} = .775$, $p = .545$, respectively.

Does the amount of target language used in the classroom, by teacher and students, have an impact on the application of oral assessment?

On the Florida Middle and High School Foreign Language Questionnaire (Appendix A), questions 29-39 addressed the amount of target language used in the classroom by teachers and students. Respondents were asked to identify whether or not the teacher or the students used the target language, rather than the base language of English, to accomplish various communicative tasks. Since these responses were "yes" or "no," the researcher identified the amount of target language used by assigning a numeric value to the number of "yes" tasks. Teachers who conducted no classroom tasks using the target language were entered as 0="none." Teachers who conducted 1, 2, 3, or 4 tasks using the target language were entered as "1," "2," "3," or "4." Examples of classroom tasks for teachers were teaching vocabulary, teaching grammar, and giving students procedural directions. The higher values indicated that the teacher used more target language in the classroom. Similarly, where students were expected by the teacher to use target language to accomplish classroom communicative tasks, such as asking to use the restroom or asking to borrow classroom supplies, the researcher entered values of "0" up to "5." The higher values suggested that more target language was used in the classroom by the students.

Nine teachers (11%) reported using no target language during the course of an instructional period, six (8%) reported using target language for two tasks, 18 (23%) reported using target language for three tasks, 29 (37%) reported using target language for four tasks, and 17 (21%) reported using target language for all five tasks. Nine teachers (11%) reported that their

students did not use any target language to accomplish instructional communicative tasks during the course of an instructional period, six (8%) reported that their students use target language to accomplish 1 task, eight (10%) reported two tasks, 15 (19%) reported three tasks, 12 (15%) reported four tasks, 19 (24%) reported five tasks and 10 teachers (13%) reported that their students used target language to accomplish six tasks in the classroom.

The dependent variable was the amount of oral assessment on unit tests. The two independent variables were the amount of target language (L2) used in the classroom by the teacher and the amount of target language (L2) used in the classroom by students. For this statistic, a univariate ANOVA was conducted. The results of the analysis are displayed in Table 11.

Table 11
Between-Subject Effects of the Amount of Target Language Used by Teachers and Students on the Administration of Oral Assessment on Unit Tests

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
	Dependent Variable: UnitAssessSpeak				
Corrected Model	16.154[a]	22	.734	1.041	.435
Intercept	87.482	1	87.482	123.974	.000
TeacherL2	.674	4	.169	.239	.915
StudentL2	6.764	6	1.127	1.598	.165
TeacherL2 * StudentL2	4.162	12	.347	.492	.911
Error	39.517	56	.706		
Total	226.000	79			
Corrected Total	55.671	78			

[a]R Squared = .290 (Adjusted R Squared = .011)
$p < .05$

91

A univariate ANOVA revealed that there was not a statistically significant difference between the means of the amount of target language used by teachers or students and the amount of oral assessment performed on unit tests: $F(12,56) = 1.04$, $p = .435$. The amount of target language used in the classroom, by teachers or students, did not have an effect on the amount of unit tests dedicated to oral assessment.

Research Question 5

If teachers do not include oral assessment in their teaching practice, what is the reason?

Hypotheses included (a) lack of instructional time, (b) lack of technological resources or technician support, (c) fear of technology, (d) test reliability issues or (e) test validity issues. On the Florida Middle and High School Foreign Language Questionnaire (Appendix A), question 43 asked respondents whether or not they included enough oral assessment in their teaching practice. The results of the analysis are displayed in Figure 6.

A total of 42 (56%) of the respondents indicated that they conducted enough oral assessment as a part of their foreign language instruction. Thirty three (44%) indicated that they did not conduct enough oral assessment as a part of their foreign language instruction. Four teachers did not answer the question. Teachers who responded "no" were asked to identify the reason that they did not include oral assessment. Of the 33 respondents who stated that they did not include enough oral assessment, 28 provided a reason.

Responses were coded into four categories: a lack of time, technology issues, class size and student resistance. Nineteen (68%) of the teachers stated that they did not feel they did enough oral assessment due to a lack of time . One teacher (4%) stated that she had frequent

technology problems. Three teachers (11%) stated that class size was the issue preventing them from doing oral assessments. These respondents did not state whether it was an issue of time or classroom management, so it was coded into a separate group. One last group of five teachers (18%) identified student resistance as the deterrent from doing oral assessments. Student resistance was the second-highest reason given for not doing oral assessments. Test reliability issues and test validity issues were not given as reasons for not including oral assessment.
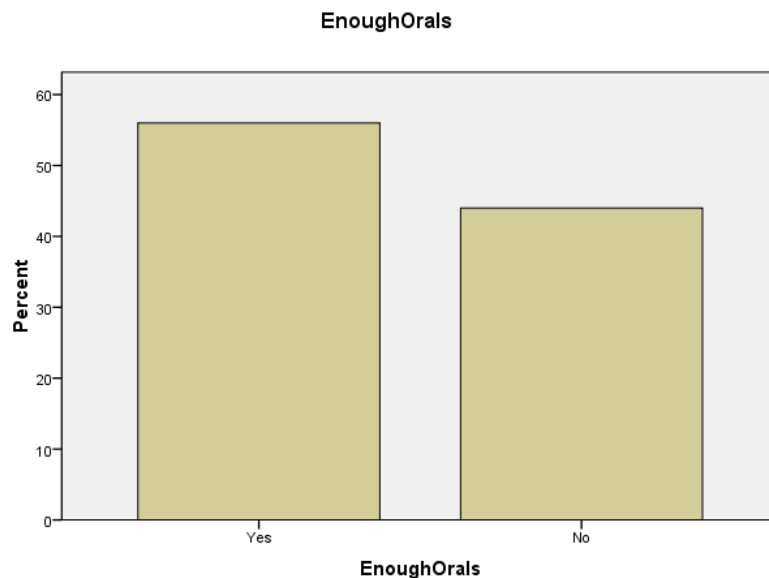

Figure 6. Teachers' opinions about whether or not they include enough oral assessment as part of instruction.

Demographic Data

Demographic data were collected at the end of the questionnaire. Respondents were asked to rate their levels of fluency in the language taught. Nine teachers (12%) self-reported

their level of fluency as "intermediate," 25 (32%) self-rated their level of fluency as "advanced,"

and 44 (56%) identified themselves as "native speakers" of the foreign language that they taught.

One teacher did not identify his or her level of fluency. The researcher performed a univariate

ANOVA. The results of the analysis are presented in Table 12.

Table 12

Between-Subjects Effects of Teacher Fluency on the Amount of Oral Assessment Administered on Unit Tests

	Dependent Variable: UnitAssessSpeak				
Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	2.523[a]	2	1.262	1.857	.163
Intercept	107.395	1	107.395	158.045	.000
FluencyLevel	2.523	2	1.262	1.857	.163
Error	50.964	75	.680		
Total	226.000	78			
Corrected Total	53.487	77			

[a]R Squared = .047 (Adjusted R Squared = .022)
p < .05

The self-reported level of fluency of the teacher was not statistically significant: $F(2,77)$

= 1.857, p = 0.163. The fluency level of the teacher did not have an effect on the amount of oral

assessment administered on unit tests. It is interesting to note that more than half of the

respondents were native speakers.

The respondents were also asked whether they taught at the middle or high school level.

Ten respondents (13%) taught at the middle school level, and 66 respondents (84%) taught at the

high school level. It would seem that more high school teachers responded to the questionnaire,

but the population to be sampled included many more high school teachers than middle school teachers. Two respondents did not answer this question. Two respondents reported teaching at both a middle school and a high school and were included in the high school group. A univariate ANOVA was conducted. The results of the analysis are displayed in Table 13.

Table 13
Between-Subjects Effects of Middle or High School on the Administration of Oral Assessment on Unit Tests

		Dependent Variable: UnitAssessSpeak			
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.002[a]	1	.002	.003	.958
Intercept	77.370	1	77.370	108.058	.000
School	.002	1	.002	.003	.958
Error	52.985	74	.716		
Total	221.000	76			
Corrected Total	52.987	75			

[a]R Squared = .000 (Adjusted R Squared = .013)
$p < .05$

Whether the respondent taught at a middle or high school was not statistically significant: $F(1,75) = .003$, $p = 0.958$. Whether the teacher taught at a middle school or a high school did not have an effect on the amount of oral assessment conducted on unit tests.

Respondents were asked to identify if their schools were Title 1 schools. At a Title 1 school, greater than 50% of the students qualify for free or reduced school lunch. This implies that the school serves an underprivileged population and qualifies for additional funding from the federal government. A total of 31 respondents (40%) reported teaching at a Title 1 school, 46

respondents (60%) reported not teaching at a Title 1 school. One respondent did not answer this

question. A univariate ANOVA was performed. Table 14 contains the results of the analysis.

Table 14

Between-Subject Effects of Community Socio-economic Status on the Administration of Oral
Assessment on Unit Tests

		Dependent Variable: UnitAssessSpeak			
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.289[a]	1	.289	.444	.507
Intercept	156.548	1	156.548	241.098	.000
Title1	.289	1	.289	.444	.507
Error	48.698	75	.649		
Total	209.000	77			
Corrected Total	48.987	76			

[a]R Squared = .006 (Adjusted R Squared = .007)
p < .05

Whether or not the respondent taught at a Title I school was not statistically significant:

$F(1,76) = 0.444$, $p = 0.507$. The socio-economic status of the population being served by the

teacher did not have an effect on the amount of unit tests dedicated to oral assessment.

Lastly, respondents were asked to identify whether they were male or female. There were

11 male respondents (14%) and 68 female respondents (86%). The numbers may be too disparate

to compare the means, but a univariate ANOVA was performed. The results of the analysis are

displayed in Table 15.

96

Table 15

Between-Subjects Effects of Gender on the Administration of Oral Assessment on Unit Tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	Dependent Variable: UnitAssessSpeak				
Corrected Model	2.803[a]	1	2.803	4.083	.047
Intercept	61.284	1	61.284	89.259	.000
Gender	2.803	1	2.803	4.083	.047
Error	52.868	77	.687		
Total	226.000	79			
Corrected Total	55.671	78			

[a]R Squared = .050 (Adjusted R Squared = .038)
p < .05

Whether or not the respondent was male was statistically significant: $F(1,78) = 4.083$ p = 0.047. Thus, the null hypothesis was rejected. There was a statistically reliable difference between the means of the two independent groups, men and women. As previously noted, however, with less than 20% of the respondents in the male group, the number may be too disparate to compare.

CHAPTER 5
DISCUSSION AND CONCLUSIONS

Summary and Discussion of Research Findings

Research Question 1

What is the actual allocation of instructional time and assessment on each of ACTFL's
five elements of foreign language proficiency?

An independent t-test revealed that the instructional time dedicated to speaking, listening,

reading, and cultures was not in alignment with the amount of unit tests dedicated to those same

skills. The amount of time dedicated to the instruction and assessment of writing was not

statistically significant, indicating that the instruction and assessment of writing were equal.

When comparing the mean responses of the alignment of each of the five skills, however, it was

revealed that, although statistically significant, the means of instruction and assessment of the

other four skills were actually somewhat different. The means for instruction and assessment of

writing were negligibly close. The mean amount of listening instruction was about half of the

time, compared to the mean amount of assessment at less than a quarter. The means for

instruction of speaking and cultures were greater by about 12% of allocated time than were the

means for the amount of assessment. Interestingly, but not surprisingly, reading was the only

skill assessed more, about 12% of the time, than instructed.

Assessments often require a lot of reading, supposedly synthesizing the material that has

been learned. Unfortunately, students who do not read well or who do not complete written

homework practice might struggle on a test, despite their daily success in the classroom with

limited reading instruction. They may speak well and comprehend the language when it is spoken and yet score poorly on a unit test with a reliance on the student's ability to comprehend the written target language. This might be even more unfair for those students who are learning a non-phonetic language, one which is not written as it is pronounced.

It was surprising to the researcher that there were not more listening comprehension on unit tests. All respondents stated that they had access to a teacher resource package. State-approved textbook adoptions typically included audio materials as well as other instructional peripheral devices. Textbook listening comprehension assessment is often offered in multiple choice formats, so it can be graded quickly and reliably. With the availability of pre-recorded audio that is tailored to each specific unit, it was surprising that the mean amounts of listening instruction and listening assessment were so disparate. Although hearing target language for "more than 50%" of instructional time, students' listening comprehension was assessed "less than 25%" of a typical unit test.

With five foreign language skills, each skill would theoretically be instructed and assessed 20% of the time. The researcher avoided offering "about 20% of the time" as a response option, forcing respondents to consider "less than 25% of the time" or "about 25% of the time" for each skill that they felt was equally instructed and assessed. With an uneven division between the 5 skills, respondents were forced to reflect on their practices rather than simply select the easiest response.

Many of the skills overlapped and the questionnaire artificially isolated them. For example, it is difficult to separate reading from writing when students are being asked to fill-in the blanks with vocabulary. Reading and writing overlap. Directions and examples are often

99

written in target language when students are instructed on speaking prompts. Reading and speaking frequently overlap, also. As questionnaire respondents considered how much time during a typical class period they allocated to reading, they had to realize that reading is often involved in the assessment of other skills. In spite of the amount of reading that is required in the instruction of the other skills, it was instructed less than assessed according to the means.

Respondents answered the question about the amount of time that students were "hearing" target language. Whether or not they were actually being instructed on listening comprehension, they were exposed to the target language from their surroundings. They were hearing their teacher, their classmates, a video or the audio practice that accompanies each of the textbook series. Dictation is a common instructional activity in the non-phonetic languages. In this activity, students write down exactly what they hear, listening to and writing in target language. It was difficult for respondents to answer this series of questions, since reading, writing, listening, speaking and cultures are often required simultaneously.

Research Question 2

Does the allocation of textbook and computer resources have an impact on the alignment of oral proficiency goals and assessments?

All respondents reported having a teacher resource package which typically includes teacher editions of textbooks, complete with standards and suggested pedagogy in the margins, workbooks, listening comprehension materials, video materials, and electronic assessment files. There was no statistical significance between the means of oral assessment conducted on unit

tests among those teachers with regular access to a computer laboratory and those without. Surprisingly, only one-third of the teachers reported having such access.

Two-thirds of the respondents reported not having "regular weekly access" to a computer laboratory. This was shocking. Computer laboratories have become common in schools, but perhaps they are utilized more frequently in remediation and credit retrieval for graduation, rather than as a means by which to enrich foreign language instruction. Computer laboratories allow students to access the world through the internet. They allow students to progress at their own pace through individual instruction. They allow teachers to collect voice data simultaneously. They motivate K-12 students, who do not know a world without technology. The possibilities of practical computer application constantly increase. Perhaps the lack of regular weekly access is due to a resistance on the part of the teachers. They might not seek to include computers as a part of instruction, intimidated by a fear of the technology or of looking foolish in front of their more technologically-advanced students. Of the respondents, 30% had 20 or more years of experience. This was the largest group, in terms of experience, among the respondents. It is possible that the age of the respondents had an effect on the amount of computer access sought.

Research Question 3

Does the level of experience of the teachers, identified as (a) number of years teaching foreign language and (b) level of education, have an impact on the alignment of goals and assessments?

The interaction between the level of experience and the level of education of the teacher, identified as having a bachelor's degree or a graduate degree, was statistically significant. The amount of time dedicated to teaching and the amount of time dedicated to study had an impact on the amount of oral assessment administered on unit tests. This significance was not surprising, as teachers with more experience and more education are often dedicated to the skill of teaching. They have continued to work in a difficult profession with a high novice attrition rate. They have developed themselves professionally by obtaining a graduate degree. They have been more likely to work to continuously improve, valuing and incorporating each of the skills into instruction, even those that are difficult to assess. Seasoned teachers with graduate degrees may have learned to manage the classroom and their time well enough that they can efficiently accomplish the oral assessment of their students.

One last point to consider is the level of language that is often taught by the more seasoned teachers. With 30% of the respondents possessing at least 20 years of experience, it is probable that they teach upper level foreign language classes. Often the teachers with less experience teach the elementary levels. Advanced students of language are motivated differently than the beginners. They enjoy the subject matter enough to continue their study. They have been successful at the prerequisite levels. They approach each of the skills more confidently, including the skill of speaking. These teachers might be preparing students for an advanced placement or

an international baccalaureate (IB) test that includes all five of the skills. These students in

particular are typically motivated to learn each of the required skills.

Research Question 4

Does the amount of target language used in the classroom by teacher and students have an impact on the application of oral assessment?

The amount of language use was measured by the number of tasks that were

accomplished in the classroom using the target language. The amount of target language used in

the classroom by teachers and students was not statistically significant. This result was surprising

not because of the lack of statistical significance but because of the lack of target language used

in some classrooms. Of the respondents, 11% stated that they did not use any target language

during a typical class period. An additional 11% of respondents reported that their students did

not use any target language in the classroom. These were not necessarily the same respondents.

Again, this research question might have been affected by the higher level of language that is

often taught by the more seasoned teachers. Advanced students can comprehend and utilize more

target language than their elementary peers. More than one-fifth of the respondents stated that

they taught using target language immersion which means that they speak only the target

language to their students.

Research Question 5

If teachers do not include enough oral assessment in their teaching practice, what is the reason?

More than half of the teachers (56%) self-reported that they administered enough oral assessment to their students. A portion of these might be the teachers of higher levels of language. Slightly less (44%) of the respondents self-reported that they did not include enough oral assessment as a part of instruction. Not surprisingly, teachers most often stated that a lack of time was the reason for not conducting what they perceived to be enough oral assessment. Oral assessment can be time-consuming, particularly in the large classes that are typical among the elementary levels. These are often the classes that are taught by teachers with less experience or less seniority on campus.

Respondents also cited too many students in a class. Too many students in a class might be considered as having a lack of time during an instructional period to listen to students' oral assessments. Too many students in a class might be a classroom management issue. When a teacher is engaged with a student in an oral assessment, the rest of the class can easily be off-task, particularly at the lower levels, where the classes are often larger and the students are younger. Teachers can be in an awkward position, hoping to encourage uncomfortable students struggling to speak in the target language while battling their own nerves. Meanwhile, teachers must stay alert to the behavior of the rest of the class for whom they are also responsible. This is no easy task, even for the seasoned teacher of more advanced students.

Student resistance was offered as a reason for not administering what the respondents considered to be enough oral assessment. Although overlooked in the hypotheses of the research

study, this was not a surprising result. Student resistance is an issue of motivation. Students in the beginner levels in particular might be threatened by their fear of speaking in the target language. They might lack the confidence that more experienced students possess. Students might be afraid of looking foolish in front of their peers. There may be a misconception that starting to learn a foreign language at middle or high school age is too old. The researcher has found this to be a common misconception among high school age children, particularly when the teacher is a native-speaker of the target language or an advanced speaker whose target language instruction started early.

Problems with technology were also given as a reason for not administering what the respondent considered to be enough oral assessment. Initially, it surprised the researcher that this was not a more common complaint from the respondents. The researcher has regular access to a computer laboratory with simultaneous voice data collection and headsets for each computer. The computers are older than the voice data software that is installed. Unreliable technology is often an issue. A computer might work one morning and malfunction in the afternoon. Technology can be a great help and a great hindrance when it does not function as intended. The lack of similar complaint in the results might be explained by the lack of regular weekly access to a computer laboratory that was shared by the majority of the respondents. Rather than complain about the unreliable but dedicated foreign language laboratory, the researcher might be better-advised to appreciate the technology that is available.

Demographic data were gathered. Whether the teacher taught at a middle or high school was not statistically significant. Of the respondents, 84% taught at a high school compared to 16% at a middle school. This is a higher response from middle school teachers than the

researcher had expected. Fewer students have been enrolled in foreign language classes at the middle school level, and it follows that there have been fewer middle school foreign language teachers. Recalling the low number of middle school programs in the districts within the population, this was a high response rate from the middle school teachers.

Whether or not the school was a Title 1 school was not statistically significant. A Title 1 school serves a lower socio-economic population. Slightly less than half of the teachers surveyed reported teaching at a Title 1 school.

The level of fluency of the teacher, self-rated as intermediate, advanced or native-speaker was not statistically significant. Interestingly, more than half of the respondents were native-speakers of the language taught. This might be due to the origin of the population of Florida's foreign language teachers. Alternatively, it might suggest that native speakers were more interested in responding to a questionnaire that concerns foreign language instruction. With so many of the questionnaire respondents being native-speakers of the language taught, these teachers might be more likely to teach the upper levels of language. The upper levels are often smaller classes that contain more motivated students with a greater breadth of target language who are less resistant to oral assessment.

The gender of the respondents caused a statistically significant difference in the mean amount of oral assessment administered on unit tests. A total of 86% of the respondents were female. Less than one-fifth, only 14%, were male. This might be too large a difference between the groups to effectively compare means. However, the difference in numbers between males and females was very interesting, and was consistent with an observation made by the researcher. At the FFLA (Florida's Foreign Language Association) conference on October 18-20, 2007, the

researcher counted 180 people at the opening ceremony, of which only 32 were males, less than a fifth of the group in attendance at the conference. More research may be warranted on this topic, which will be addressed in a later section.

Discussion of the Research Process

The random sample of the population included 175 teachers. It should be noted that 79 teachers consented to respond to the questionnaire which was 45% of the sample. If the IRB process had been started earlier in the K-12 school year, it would have allowed more time for follow up with teacher-participants. Additionally, two participants noted that the questionnaire being sent earlier in the school year might have been more convenient for teachers, since the end-of-year with its many deadlines is a hectic time for teachers.

The researcher was pleased with the decision to use "Quia," software that was familiar to teachers rather than more commonly-used online research software such as "Survey Monkey." Although the latter offers a more visually-appealing online survey tool, the former is familiar to this particular group of respondents. Quia is quite popular among foreign language teachers. As a classroom teacher, the researcher has maintained a subscription to Quia for use with computer-assisted instruction, practice, online assessment, and student survey. The researcher was already familiar with the software, and it made no sense to pay for a subscription to an additional online tool.

It would have been more effective to define the term " L2" on the questionnaire. L2 refers to second language. In an effort to abbreviate the time required to complete the questionnaire, the researcher made an assumption that all foreign language teachers are familiar

with this abbreviation for second language. Several questionnaire respondents asked the question via email prior to completing the questionnaire. Others stated as a comment on the completed questionnaire that they were unfamiliar with the L2 abbreviation. Several colleagues asked if it meant Level 2 language which it does not. It was surprising to the researcher that several seasoned and masters-degreed colleagues were unfamiliar with this expression which is commonly-used at conferences, in the literature, and in text related to second language acquisition. The researcher does not believe that the research outcomes were affected by the few respondents who completed the questionnaire and then indicated an ignorance of the L2 expression.

One respondent, a Latin teacher, suggested that not all languages are living, spoken languages and that respondents should have had the opportunity to identify the language taught, as a justification for not incorporating oral instruction and assessment. Despite the continued use of spoken Latin in some circumstances, this is a valid point. Latin is certainly less-widely spoken than the living languages. Teachers of the classical languages should probably have been removed from the population.

Survey question 4 asked "During a typical class period, how much time are students hearing L2?" It is possible that respondents understood this question to be a question of immersion, rather than a question of instructional time dedicated to listening comprehension practice. It would have been more-clearly understood if the researcher had more specifically stated "doing listening practice."

Survey question 27 asked respondents whether or not they had regular (weekly) access to a computer laboratory. Survey question 28 asked respondents if that computer laboratory had

headsets/microphones. More respondents answered "yes" to question 28 than 27, implying that there was more access to computers with headsets and microphones than just computers. This question should have been more clearly written, in order to obtain better data.

In self-reporting, good teachers are often hard on themselves, always comparing themselves to an ideal goal. Respondents were asked, in survey question 43, whether or not they did enough oral assessment. The 33 respondents who stated that they did not do enough oral assessment might still do more assessment than other teachers. They may even have done more oral assessment than the teachers who self-rated as doing enough.

Similarly, survey question 42 asked respondents to self-rate their fluency. With an awareness of the ACTFL oral proficiency scale, many teachers can accurately rate themselves. Without such awareness, non-native speakers might underestimate their fluency and rate themselves lower than their actual ability, particularly when comparing themselves to the native speakers that are prevalent in the field of foreign language education.

With a response rate of 45% to an online questionnaire, it seems likely that respondents were the sort of professionals who would be involved in the research process and who have a certain level of dedication to their profession of teaching foreign language. These respondents will likely be teaching their students to the best of their ability within their situations. They will be mindful of continuous improvement and involved in professional development. The non-respondents may have caused the results to be different. The outcomes of this research study could be biased due to the level of dedication of the respondents who responded to a colleague's online questionnaire.

Implications of the Findings

When comparing the means of instruction versus assessment of each of the five skills of foreign language instruction, writing was the only skill that was not statistically significant. The means of instruction and assessment of writing were equal, implying that instruction and assessment of writing were in alignment. All other skills were not equal, meaning that instruction and assessment were not aligned. Despite being statistically significant, the means of instruction versus assessment in reading, speaking, and cultures were close, suggesting that goals and assessment of those skills were not far out of alignment. Listening, on the other hand, although significant, appeared to be instructed much more than assessed. It is not surprising to note that reading was the only skill that was assessed more than instructed, presumably due to ease of assessment.

There was an interaction between the level of education and experience of the teacher as having an effect on the inclusion of oral assessment on unit tests. This alignment of goals and assessment might be affected by the more advanced level of student that seasoned and vested teachers often teach. The seasoned teacher with a graduate degree might be more likely to participate in research. The gender of the instructor was also significant although difficult to compare the disparate means that were revealed. More interesting was that one-fifth of the respondents were male and four-fifths were female. There appear to be many more female than male foreign language educators at the middle and high school levels.

More than half of the respondents were native speakers, which might suggest that native speakers were more likely to align instruction and assessment of the five skills according to the

standards. It might also mean that they were more interested in participating in the research or that they reflected the population of Florida's foreign language teachers.

Two-thirds of the respondents stated that they did not have regular weekly access to a computer laboratory for their students. This was a surprising finding that might be due to administrative allocation of computer laboratories to other subject areas. There might be a lack of teacher interest.

The findings of this study indicated that the alignment of goals and assessment as a part of foreign language instruction had more to do with individual teachers than with any variable that they might share, other than the interaction of years of experience and level of education. Among the teachers who felt that they did not conduct sufficient oral assessment, a lack of time, student resistance to oral assessment, class size, and technology problems were the reasons given.

<div align="center">Recommendations for Future Practice</div>

In order to align the goals for foreign language instruction, the results of this research would suggest that teachers allocate the same amount of instructional time for each of the five foreign language skills to the assessment of the same five skills. Rather than instructing listening comprehension "about 50%" of the class period and having "less than 25%" of the test involve a listening comprehension assessment, foreign language teachers might work to assess as was instructed. Similarly, although not different by more than 12% of the allocated time, reading was assessed more than it was instructed. Teachers might focus more on the alignment of instruction

and assessment than on the ease of assessment. The researcher hastens to add that she will strive to do likewise.

<u>Recommendations for Future Research</u>

Several questions were raised by the research process itself. One question that has been raised as a result of this study is how to obtain a more representative sample of the population. At the second contact, with questionnaire embedded in the email, there were about 25 responses over three days. The additional responses, up to the 79 total respondents, required third, fourth and fifth contacts. The end of the school year forced the rapid succession of research contacts and closure of the research period. It would be interesting to learn how to reach those who do not respond to questionnaires. Was an online questionnaire exclusive to those within the sample who are comfortable using technology? Perhaps the teachers who participated in this research were those teachers who were less overwhelmed and perhaps already more mindful of aligning goals and assessment. These research outcomes might have been more significant had there been a response that more accurately reflected the population.

Another question that was raised as a result of this study was the optimum timing of the questionnaire. Would the response rate have been greater than 45% if the series of contacts had occurred at a different time of year, perhaps during pre-planning or at least apart from midterm or final grades? Was any one group more likely to respond than another? The researcher did not collect any demographic data from district web sites as she was compiling her population list. Another study that includes such population demographics might indicate which groups are more likely to respond, such as native speakers or middle school teachers.

It is interesting to note that, of the 79 respondents, slightly more than half were native-speakers of the language taught. This might have had an impact on the 45% of the population sample who responded to the questionnaire. Native speakers might be more likely to respond to a questionnaire. This, in turn, could cause bias in the research outcomes, as native-speakers often teach the higher levels of language. The state of Florida has a large population of Spanish-speakers which might have contributed to the high percentage of native-speakers among the sample.

Similarly, middle school teachers represented more of the respondents than the actual population would suggest. It would be interesting to go back to the population and identify the middle school and high school teachers, so that the proportions of the sample and their responses could be compared. Were the middle school teachers really overrepresented in these results? Perhaps middle school teachers were more likely to help and support a colleague who is conducting a research study.

The large difference between the numbers of males compared to females who responded to this questionnaire raised some questions for further study. It would be interesting to know the genders of the total sample, in addition to the genders of the respondents. This would indicate whether or not males or females were more likely to participate in a research study. Is the disproportion that was evidenced among the respondents of the research study, as well as the researcher's experience at the state foreign language conference, limited to foreign language educators at the middle and high school level? Is there such disproportion among university instructors? The researcher observed a similar phenomenon during the 16 courses of her undergraduate French degree. Several students would become instructors, while others were

studying French for international business, government service or interpretation and translation. The population was noticeably female, despite the variety of professional linguistic outcomes. Perhaps that disproportion was tied to the language being studied. The large female group might not be limited to foreign language instructors but rather to the study of foreign language itself.

In future research, it would be helpful to know which languages the respondents instruct. Perhaps the instructors of one language are more likely to respond or are more likely to be male. Perhaps the instructors of one language are more likely to conduct oral assessment than another.

Conducting a research study that involved colleagues from around the state of Florida has brought the researcher closer to them, and motivates her on a new level to do the best that she can for the students in her own high school classroom. This experience has been, for her, a great combination of research and practice in the field. She plans to follow up with participants of the study, sharing with them some abbreviated outcomes of this research. The researcher hopes to encourage and motivate them, as they have encouraged and motivated her.

APPENDIX A
STANDARDS FOR FOREIGN LANGUAGE LEARNING

COMMUNICATION
Communicate in Languages Other Than English

- **Standard 1.1:** Students engage in conversations, provide and obtain information, express feelings and emotions, and exchange opinions
- **Standard 1.2:** Students understand and interpret written and spoken language on a variety of topics
- **Standard 1.3:** Students present information, concepts, and ideas to an audience of listeners or readers on a variety of topics.

CULTURES
Gain Knowledge and Understanding of Other Cultures

- **Standard 2.1:** Students demonstrate an understanding of the relationship between the practices and perspectives of the culture studied
- **Standard 2.2:** Students demonstrate an understanding of the relationship between the products and perspectives of the culture studied

CONNECTIONS
Connect with Other Disciplines and Acquire Information

- **Standard 3.1:** Students reinforce and further their knowledge of other disciplines through the foreign language
- **Standard 3.2:** Students acquire information and recognize the distinctive viewpoints that are only available through the foreign language and its cultures

COMPARISONS
Develop Insight into the Nature of Language and Culture

- **Standard 4.1:** Students demonstrate understanding of the nature of language through comparisons of the language studied and their own
- **Standard 4.2:** Students demonstrate understanding of the concept of culture through comparisons of the cultures studied and their own.

COMMUNITIES
Participate in Multilingual Communities at Home & Around the World

- **Standard 5.1:** Students use the language both within and beyond the school setting
- **Standard 5.2:** Students show evidence of becoming life-long learners by using the language for personal enjoyment and enrichment.

APPENDIX B
STUDENT ENROLLMENT BY DISTRICT

Student Membership by Grade, 2007-08

Enrollment	DISTRICT NAME	MEMBERSHIP	Enrollment	DISTRICT NAME	MEMBERSHIP
1	DADE	348,113	41	JACKSON	7,363
2	**BROWARD**	**258,895**	42	HENDRY	7,308
3	HILLSBOROUGH	193,116	43	OKEECHOBEE	7,037
4	ORANGE	174,136	**44**	**WALTON**	**6,967**
5	PALM BEACH	170,844	45	GADSDEN	6,516
6	DUVAL	124,775	46	LEVY	6,228
7	**PINELLAS**	**107,895**	47	SUWANNEE	6,005
8	POLK	94,164	48	WAKULLA	5,178
9	LEE	80,541	**49**	**HARDEE**	**5,014**
10	BREVARD	74,371	50	DESOTO	5,012
11	**PASCO**	**66,313**	51	BAKER	4,958
12	SEMINOLE	65,355	52	WASHINGTON	3,590
13	**VOLUSIA**	**64,570**	53	BRADFORD	3,576
14	**OSCEOLA**	**52,742**	54	**HOLMES**	**3,430**
15	COLLIER	42,721	55	**TAYLOR**	**3,389**
16	**MARION**	**42,565**	56	**GILCHRIST**	**2,889**
17	MANATEE	42,524	57	MADISON	2,783
18	SARASOTA	42,013	58	UNION	2,296
19	ESCAMBIA	41,855	59	CALHOUN	2,229
20	LAKE	40,710	60	*FSU LAB SCH*	*2,210*
21	**ST. LUCIE**	**40,347**	61	DIXIE	2,190
22	CLAY	36,125	**62**	**GULF**	**2,171**
23	LEON	32,471	63	HAMILTON	2,018
24	OKALOOSA	29,568	**64**	**LIBERTY**	**1,513**
25	ALACHUA	28,373	65	**GLADES**	**1,365**
26	**ST. JOHNS**	**27,867**	66	**FRANKLIN**	**1,246**
27	BAY	26,236	67	*UF LAB SCH*	*1,162*
28	SANTA ROSA	25,711	68	**JEFFERSON**	**1,154**
29	HERNANDO	22,836	69	**LAFAYETTE**	**1,089**
30	MARTIN	18,109	70	*FLVA*	*945*
31	**CHARLOTTE**	**17,799**	71	*CONNECTIONS*	*901*
32	INDIAN RIVER	17,646	72	*DEAF/BLIND*	*703*
33	CITRUS	16,174	73	*FAU LAB SCH*	*640*
34	**FLAGLER**	**12,774**	74	*FAMU LAB SCH*	*370*
35	HIGHLANDS	12,445	75	*DOZIER/OKEEC*	*331*
36	PUTNAM	11,808	**99**	**STATE**	**2,653,377**
37	*NASSAU*	*11,079*			
38	**COLUMBIA**	**10,134**			
39	MONROE	8,363			
40	**SUMTER**	**7,518**			

APPENDIX C
FOREIGN LANGUAGE ORAL ASSESSMENT PRACTICES
IN FLORIDA MIDDLE AND HIGH SCHOOLS

Please tick just one box per row. If you teach more than one language, please answer these questions with respect to the language you are most comfortable teaching.

During a typical class period, how much time are students:

	None	Less than a quarter (<25%)	Around a quarter (25%)	About half (50%)	Around three quarters (75%)	More than three quarters (>75%)
1. reading L2?	☐	☐	☐	☐	☐	☐
2. writing L2?	☐	☐	☐	☐	☐	☐
3. speaking L2?	☐	☐	☐	☐	☐	☐
4. hearing L2?	☐	☐	☐	☐	☐	☐
5. learning about cultures of L2?	☐	☐	☐	☐	☐	☐

On a typical unit test, on what percentage of the questions are your students being assessed on:

	None	Less than a quarter (<25%)	Around a quarter (25%)	About half (50%)	Around three quarters (75%)	More than three quarters (>75%)
6. reading L2?	☐	☐	☐	☐	☐	☐
7. writing L2?	☐	☐	☐	☐	☐	☐
8. listening to L2?	☐	☐	☐	☐	☐	☐
9. speaking L2?	☐	☐	☐	☐	☐	☐
10. knowledge of cultures of L2?	☐	☐	☐	☐	☐	☐

Yes   No

11.	I assess my students' reading comprehension	☐	☐
12.	I assess my students' language knowledge by having them define terms	☐	☐
13.	I assess my students' writing with short answers or fill-ins	☐	☐
14.	I assess my students' writing with essays or paragraphs	☐	☐
15.	I assess my students' listening comprehension with textbook CD's and cassettes	☐	☐
16.	I assess my students' listening comprehension by speaking to them	☐	☐
17.	I assess my students' speaking skills with dialogues or skits/plays	☐	☐
18.	I assess my students' speaking skills by presenting them with picture prompts	☐	☐
19.	I assess my students' speaking skills by presenting them with written prompts	☐	☐
20.	I assess my students' language skills by having them maintain a portfolio that includes writing AND speaking samples	☐	☐
21.	I assess my students' speaking skills with a classroom observational checklist	☐	☐
22.	I assess my students' speaking skills with student self-assessment	☐	☐
23.	I assess my students' speaking skills with oral interviews	☐	☐
24.	I assess my students' speaking skills with oral presentations	☐	☐
25.	I assess my students' speaking skills with computer-assisted voice data collection	☐	☐
		Yes	No
26.	Do you have a textbook teacher resource package?	☐	☐

27.	Do you have regular (weekly) access to a computer laboratory?	☐	☐
28.	Does the computer laboratory have headsets/microphones?	☐	☐

		Yes	No
29.	Do you use L2 in the classroom to give classroom directions?	☐	☐
30.	Do you use L2 in the classroom to give speaking/listening practice?	☐	☐
31.	Do you use L2 in the classroom to teach vocabulary?	☐	☐
32.	Do you use L2 in the classroom to teach grammar?	☐	☐
33.	Do you use L2 in the classroom to communicate all tasks--students are immersed?	☐	☐

		Yes	No
34.	Do your students use L2 in the classroom to ask for restroom, locker, water, etc.?	☐	☐
35.	Do your students use L2 in the classroom to ask to borrow classroom supplies?	☐	☐
36.	Do your students use L2 in the classroom to ask for repetition?	☐	☐
37.	Do your students use L2 in the classroom to ask for information?	☐	☐
38.	Do your students use L2 in the classroom to discuss/converse on a topic?	☐	☐
39.	Do your students use L2 in the classroom to communicate all tasks--students are immersed?	☐	☐

40. What is the highest degree that you have attained?

Bachelor degree ☐

Master's degree ☐

Specialist degree ☐

Doctoral degree ☐

41. How many years have you been teaching?

1-3 years ☐

4-6 years ☐

7-10 years ☐

11-20 years ☐

more than 20 years ☐

42. How well do you speak the language you teach?

Intermediate ☐

Advanced ☐

Native-speaker ☐

43
. Do you believe that you provide enough instruction and assessment of spoken language?
If yes, please type "yes." If no, why not? Specific reasons would be very much appreciated.

_____
___
_____
___
_____
___
_____
___

44 Do you teach at a middle school or a high
. school?

Middle school ☐

High school ☐

45
. Is your school a Title 1 school?

Yes ☐

No ☐

46
. What is your gender?

Male ☐

Female ☐

Thank you very much for your time. Enjoy the remainder of your school year.

APPENDIX D
IRB APPROVAL LETTER

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
          FWA00000351, IRB00001138

To: **Grace Kellermeier**

Date: **April 30, 2010**

Dear Researcher:

On 4/30/2010, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Foreign Language Oral Assessment Practices in Florida Middle and High Schools
Investigator:	Grace Kellermeier
IRB Number:	SBE-10-06905
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM, UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 04/30/2010 01:10:06 PM EDT

IRB Coordinator

Page 1 of 1

126

REFERENCES

Adams, R. J., Griffin, P. E., & Martin, L. (1987). A latent trait method for measuring a dimension in second language proficiency. *Language Testing, 4*(1), 9-27.

Anderson, R., Spiro, R., & Anderson, M. (1978). Schemata as scaffolding for the representation of information in connected discourse. *American Education Research Journal, 15*, 433-440.

Asher, J. (1977). Children learning another language: A developmental hypothesis. *Child Development, 48*(3), 1040-1048.

Asher, J. (1993). Imagination in second language acquisition. *Journal of the Imagination in Language Learning, 1*, 20-23.

Asher, J., Kusado, J., & de la Torre, R. (1974). Learning a second language through commands: The second field test. *The Modern Language Journal, 58* (1-2), 24-32.

Bang, Y. (2000). Incorporating the affective variables into second and foreign language curricula design. Theoretical and practical considerations. *Journal of Pan-Pacific Association of Applied Linguistics, 4*(1), 153-174.

Berrier, A. (1989). Oral evaluation: Highs and lows of interview and role playing. [Evaluation de l'oral: les hauts et les bas de l'entrevue et du jeu de role] *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 45*(2), 345-356.

Bialystock, E. (2002). On the reliability of robustness: A reply to DeKeyser. *Studies in Second Language Acquisition, 41*, 481-488.

Bialystok, E. (1998). Coming of age in applied linguistics. *Language Learning, 48*, 497-518.

Bialystok, E., & Miller, B. (1999). The problem of age in second language acquisition:

Influences from language, task and structure. *Bilingualism: Language and Cognition, 2*, 127-145.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*, 706-755.

Blake, R., Wilson, N. L., Cetto, M., & Pardo-Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning & Technology, 12*(3), 114- 127.

Bonk, W. J., & Ockey, G. J. (2003). A many-facet rasch analysis of the second language group oral discussion task. *Language Testing, 20*(1), 89-110.

Borras, I. (1993). Developing and assessing practicing spoken French: A multimedia program for improving speaking skills. *Educational Technology Research and Development, 41*(4), 91-103.

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT and subsequent performance. *System, 34*, 15-35.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development* . Boston, MA: Heinle & Heinle

Bruhn, T. C. (1989). 'Passages': Life, the universe, and language proficiency assessment. *Georgetown University Round Table on Languages and Linguistics, 1989.*

Brumfit, C. (1983). Language variation and the death of language teaching. *Jyvaskyla cross-language series, 9*, 125-132.

Brumfit, C. (1995). Theoretical practice: Applied linguistics as pure and practical science. *AILA Review, 12*, 18-30.

Burnett, J. (1998). Language alternation in a computer-equipped foreign language classroom; The intersection of teacher beliefs, language and technology. *The Canadian Modern Language Review, 55*(1), 97-123.

Cafarella, C. (1997). Assessor accommodation in the V.C.E.Italian oral test. *Australian Review of Applied Linguistics, 20*(1), 21-41.

Capretz, P. (1979). Une experience d'utilisation du film. *Le Français Dans le Monde, 19*(145), 90-94.

Capretz, P. (1989). A video-based system for the teaching of French language and culture. *The Journal of Educational Techniques and Technologies, 22*(2-3), 2-7.

Center for Applied Linguistics Online Resources: Digests. (1995). *Guidelines for starting an elementary school foreign language program.* Retrieved June 16, 2009, from http://www.cal.org/resources/digest/rosenb01.html

Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tests and rater groups. *Language Testing, 12*(1), 16-33.

Chambers, F. (1997). What do we mean by fluency? *System, 25*(4), 535-544.

Chambers, F., & Richards, B. (1992). Criteria for oral assessment. *Language Learning Journal, 6*(Sept), 5-9.

Chambers, F., & Richards, B. (1993). Oral assessment: The views of language teachers. *Language Learning Journal, 7*(Mar), 22-26.

Chambers, F., & Richards, B. (1995). The 'free conversation' and the assessment of oral proficiency. *Language Learning Journal, 11*(Mar), 6-10.

Chiu, T., Liou, H., & Yeh, Y. (2007). A study of web-based oral activities enhanced by

    automatic speech recognition for EFL college learning. *Computer Assisted Language*

    Learning, 20(3), 209-233.

Choi, J. Y. (2006). Metacognitive evaluation method in consecutive interpretation for novice

    learners. *Meta, 51*(2), 273-283.

Clark, J. (1981). Communication in the classroom. *Modern Languages in Scotland, 21-   22*,

    144-156.

Clark, J. (1981). A communicative approach in a school setting. [Une approche

    communicative dans un context scolaire] *Le Français Dans Le Monde, 21*(160), 29-38.

Clark, J., & Hooshmand, D. (1992). "Screen-to-screen" testing: An exploratory study of   oral

    proficiency interviewing using video teleconferencing. *System, 20*(3), 293-304.

Clifford, R. (1989). Technological, methodological, and assessment challenges: Can the   foreign

    language teacher survive? *Georgetown University Round Table on Languages and*

    Linguistics, 198-205.

Clifford, R., ed. (2003). Oral proficiency testing. *Foreign Language Annals, 36*(4), 481-588.

Cummings, A. (2005). Administrative and pedagogical uses of computers in foreign

    language classrooms: A survey of Spanish teachers' beliefs and practices. (PhD, The

    University of Iowa).

Curtain, H. & Dahlberg, C. (2004). *Languages and children--making the match: New languages*

    for young learners. Boston, MA: Pearson Education, Inc.

Davies, S., Irvine, A., & Larrieu, J. (1997). Oral classroom testing in an adult French community

    class. *Edinburgh Working Papers in Applied Linguistics, 8,* 24-43.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher

assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305-334.

Davison, C. (2007). Views from the chalk face: English language school-based assessment in

Hong Kong. *Language Assessment Quarterly, 4*(1), 37-68.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition.

Studies in Second Language Acquisition, 22, 499-533.

Deceiver, R. (1993). The effect of error correction on L2 grammar knowledge and oral

proficiency. *The Modern Language Journal, 77*(4), 501-514.

Dewaele, J. (2005). Sociodemographic, psychological and politicocultural correlates in Flemish

students' attitudes towards French and English. *Journal of Multilingual and Multicultural Development, 26*(2), 118-137.

Dillman, D. (2000). *Mail and internet survey*s. New York: John Wiley & Sons, Inc.

Doughty, C., & Long, M. (2003). *Handbook of second language acquisition*. Oxford:

Blackwell Publishing.

Douglas, D. (2001). Language for specific purposes assessment criteria: Where do they come

from? *Language Testing, 18*(2), 171-185.

Early, P., & Swanson, P. (2008). Technology for oral assessment. *Dimension,* 39-47.

Ehrman, M., Leaver, B., & Oxford, R. (2003). A brief overview of individual differences in

second language learning. *System, 31*(3), 313-330.

Ellis, R. (1984). Communication strategies and the evaluation of communicative performance.

English Language Teaching Journal, 38(1), 39-44.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Fall, T., Adair-Hauck, B., & Glisan, E. (2007). Assessing students' oral proficiency: A case for online testing. *Foreign Language Annals, 40*(3), 377-400.

Farber, B. (1991). *How to learn any language quickly, easily, inexpensively, enjoyably and on your own.* New York: Carol Publishing Group (Citadel Press)

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Frink, H. (1982). Oral testing for first-year language classes. *Foreign Language Annals, 15*(4), 281-287.

Gan, Z. (2008). Examining negotiation in peer group oral assessment. What are the implications? *Australian Review of Applied Linguistics, 31*(1), [np].

Garcia, P. & Hernandez, T. (2007). Information gap activities: A standards-based strategy for promoting oral proficiency in a thematic context. *Florida Foreign Language Journal, 4*(1), 10-17.

Garcia-Laborda, J. (2004). HIEO: Research and development in a computer based tool of multilingual oral evaluation. [HIEO: Investigacion y desarrollo de una herramienta informatica de evaluacion oral multilingue] *Didactica (Lengua y Literatura), 16*, 77-88.

Gasparro-Liskin, J. (1985). The ACTFL proficiency guidelines: A historical perspective. ED238265, retrieved 06/29/06 from http://eric.ed.gov

Govoni, J., & Feyten, C. (1999). Effects of the ACTFL-OPI-type training on student performance, instructional methods, and classroom materials in the secondary foreign language classroom. *Foreign Language Annals, 32*(2), 189-204.

Greenhough, T. (1994). Criterion-referenced assessment for Japanese-the Welsh experience. *Language Learning Journal, 10*(Sept), 82-83.

Hahn, S. L., Stassen, T., & Reschke, C. (1989). Grading classroom oral activities: Effects on motivation and proficiency. *Foreign Language Annals, 22*(3), 241-252.

Hallam, M. (2008). Language laboratories: A 21st century tool for 21st century learning. *The Language Educator, 3*(6), 34-37.

Halleck, G. (2007). Data generation through role-play: Assessing oral proficiency. *Simulation & Gaming, 38*(1), 91-106.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*(3), 379-400.

Henning, G. (1983). Oral proficiency testing: Comparative validities of interview, imitation, and completion methods. *Language Learning, 33*(3), 315-332.

Henning, G. (1992). The ACTFL oral proficiency interview: Validity evidence. *System, 20*(3), 365-372.

Hernandez, T. (2006). Integrative motivation as a predictor of success in the intermediate foreign language classroom. *Foreign Language Annals, 39*(4), 605-617.

Higgs, T. V. (1987). Oral proficiency testing and its significance for practice. *Theory into Practice, 26*(4), 282-287.

Hoekje, B., & Linnell, K. (1994). "Authenticity" in language testing: Evaluating spoken language tests for international teaching assistants. *TESOL Quarterly, 28*(1), 103-126.

Horwitz, E., Horwitz, M., & Lope, J. (1986). Foreign Language Classroom Anxiety. *Modern Language Journal, 70*(2), 125-133.

Hughes, R. (2004). Testing the visible: Literate biases in oral language testing. *Journal of Applied Linguistics, 1*(3), 295-309.

Hurman, J. (1993). Assessing the prepared topic in A-level speaking. *Language Learning Journal, 7*(Mar), 9-12.

Ikeda, K. (1998). The paired learner interview: A preliminary investigation applying vygotskian insights. *Language, Culture and Curriculum, 11*(1), 71-96.

Im, B. (2000). Understanding and application of performance assessment. *Journal of Pan-Pacific Association of Applied Linguistics, 4*(1), 41-72.

Ivers, J. (1990). A freirian approach to world view instruction in the foreign language curriculum. (PhD. University of North Carolina at Greensboro).

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51*(3), 401-436.

Jacoby, L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36*, 300-324.

Jafarpur, A. (1996). Native speaker performance validity: In vain or for gain? *System, 24*(1), 83-95.

James, C., Bowsher, J., & Simpson, P. (1991). Speech idiometry: Digitization effects and the non-equivalence of isophonemic word lists. *British Society of Audiology, 25*, 111-121.

Jenkins, S., & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal, 87*(1), 90-107.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Johnson, M. (2000). Interaction in the oral proficiency interview: Problems of validity. *Pragmatics, 10*(2), 215-231.

Jung, H. (2000). The effectiveness of measures of general and oral language proficiency, academic achievement, motivation, and attitudes in the selection of Korean military candidates for overseas training. (Ph.D., The Pennsylvania State University).

Kasper, G., & Ross, S. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics, 39*(11), 2045-2070.

Ke, C. (1993). An empirical investigation of the relationship between a simulated oral proficiency interview and the ACTFL oral proficiency interview. *Selecta, 14*, 6-10.

Kenyon, D., & Tschirner, E. (2000). The rating of direct and semi-direct oral proficiency interviews: Comparing performance at lower proficiency levels. *The Modern Language Journal, 84*(1), 85-101.

Kim, H. (1980). Language proficiency testing with special reference to oral English. [(Title in Korean; see translation)] *Ohak Yonku/Language Research, 16*(2), 197-224.

Kozaki, Y. (2004). Using GENOVA and FACETS to set multiple standards on performance

    assessment for certification in medical translation from Japanese into English. *Language*

    Testing, 21(1), 1-27.

Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language*

    Journal, 70(4), 366-372.

Krashen, S. (1987). *Principles and practice in second language acquisition*. Prentice-Hall

    International.

Kugler-Euerle, G. (1992). Performance evaluation in English instruction at the middle level.

    [Leistungsmessung im Englischunterricht der mittelstufe] *Fremdsprachenunterricht,*

    36(5), 263-266.

Kuhl, P. (1993). Early linguistic experience and phonetic perception: Implications for theories

    of developmental speech perception. *Journal of Phonetics, 21*(1-2), 125-139.

Kuo, J., & Jiang, X. (1997). Assessing the assessments: The OPI and the SOPI. *Foreign*

    Language Annals, 30(4), 503-512.

Lazaraton, A., & Davis, L. (2008). A microanalytic perspective on discourse, proficiency, and

    identity in paired oral assessment. *Language Assessment Quarterly, 5*(4), 313-335.

Lazaraton, A., & Riggenbach, H. (1990). Oral skills testing: A rhetorical task approach. *Issues in*

    Applied Linguistics, 1(2), 196-217.

Lee, Y. (2007). The multimedia assisted test of English speaking: The SOPI approach. *Language*

    Assessment Quarterly, 4(4), 352-366.

Lightbrown, P. & Spada, N. (2006). *How Languages are Learned*. Oxford: Oxford University

    Press.

Luecht, R. (2003). Multistage complexity in language proficiency assessment: A framework

    for aligning theoretical perspectives, test development, and psychometrics. *Foreign*

    Language Annals, 36(4), 527-535.

Lumley, T., & McNamara, T. (1995). Rater characteristics and rater bias: Implications for

    training. *Language Testing, 12*(1), 54-71.

Magnan, S. S. (1991). Just do it: Directing TAs toward task-based and process-oriented testing.

    In R. V. Teschner (Ed.), *Assessing foreign language proficiency of undergraduates* (pp.

    135-161). Boston, MA: Heinle & Heinle.

Malone, M., Rifkin, B, Christian, D., & Johnson, D. *Attaining high levels of proficiency:*

    Challenges for foreign language education in the United States. Center for Applied

    Linguistics. Retrieved March 18, 2009 from http://w.cal.org/resources/digest/attain.html

Manley, J. (1995). Assessing students' oral language: One school district's response. *Foreign*

    Language Annals, 28(1), 93-102.

Marshall, J., & Peters, M. (1989). Te reo o te tai tokerau: The assessment of oral Maori. *Journal*

    of Multilingual and Multicultural Development, 10(6), 499-514.

Mathews, T., & Hansen, C. (2004). Ongoing assessment of a university foreign language

    program. *Foreign Language Annals, 37*(4), 630-640.

McAlpine, D., & Dhonau, S. (2007). Creating a culture for the preparation of an

    ACTFL/NCATE program review. *Foreign Language Annals, 40*(2), 247-259.

McCarthy, P. (1978). *The teaching of pronunciation*. Cambridge: Cambridge University Press.

McColl, H. (2005). Foreign language learning and inclusion: Who? Why? What? *Support for*

    Learning, *20*(3), 103-108.

McMillan, J. & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49.

McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18*(4), 446-466.

Mentz, O. (2002). Changes in French instruction need a different form of evaluation for students' performances. [Ein veranderter Franzosischunterricht braucht eine andere form der beurteilung von schulerleistungen] *Neusprachliche Mitteilungen Aus Wissenschaft Und Praxis, 55*(3), 134-141.

Meredith, R. (1990). The oral proficiency interview in real life: Sharpening the scale. *The Modern Language Journal, 74*(3), 288-296.

Mikhailova, J. (2007). Rethinking description in the Russian SOPI: Shortcomings of the simulated oral proficiency interview. *Foreign Language Annals, 40*(4), 584-603.

Mir, M. (2006). Using "oral journals" to develop speaking fluency, self-confidence and much more! *Hispania, 89*(3), 559-561.

Mitchell, R., & Brumfit, C. (1997). The national curriculum experience of bilingual students. *Educational Review, 49*, 159-180.

Moeller, A., Reschke, C., & Hahn, S. (1993). Graded or nongraded communicative activities-what is the answer?-an update. *Foreign Language Annals, 26*(1), 75-80.

Nakatani, Y. (2006). Developing an oral communication strategy inventory. *The Modern Language Journal, 90*(2), 151-168.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

138

Nibungco, J., & Williams, M. (1996). Designing oral assessment for nontraditional ESL

    students in a community college. *College ESL, 6*(1), 85-94.

Nieman, M. (1997). The implications of competence-based teacher education for the training

    of second language teachers. [Die implikasies van bevoegdheidsgebaseerde

    onderwysersopleiding vir die opleiding van tweedetaalonderwysers] Tydskrif vir

    Taalonderrig/*Journal for Language Teaching*, *31*(2), 98-113.

Nikolov, M., & Krashen, S. (1998). Need we sacrifice accuracy for fluency? *System, 25*(2),

    197-201.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and

    quantitative analysis. *Language Learning, 50*, 417-528.

Norrish, N. (1978). Using the language laboratory to test linguistic performance and

    communicative ability. *System, 6*(1), 30-40.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, England:

    Cambridge University Press.

Nutta, J., Feyten, C, Norwood, A., & others. (2002). Exploring new frontiers: What do

    computers contribute to teaching foreign languages in elementary school? *Foreign

    Language Annals*, *35(*3), 293-306.

Olaofe, I. A. (1992). A communicative model for assessing second language performance. *IRAL,

    30*(3), 207-222.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing,

    19*(2), 169-192.

O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing, 19*(3), 277-295.

Othman, J., & Poh, C. (2005). Cooperative learning instruction: Its impact on ESL learners' oral performance. *Journal of Pan-Pacific Association of Applied Linguistics, 9*(1), 37-58.

Oxford, R. & Ehrman, M. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13*, 188-205.

Ozerk, K., & Krashen, S. (2001). Subject matter teaching in bilingual education: Impact on bilingual and monolingual students. *ITL, Review of Applied Linguistics*, *131-132*, 1-10.

Padilla, A., Aninao, J., & Sung, H. (1996). Development and implementation of student portfolios in foreign language programs. *Foreign Language Annals, 29*, 429-438.

Paulus, L. (1998). Watch them SOAR: Student oral assessment redefined. *Hispania, 81*(1), 146-152.

Price, E. (1980). Monitoring attainment in the Welsh language in schools. *Journal of Multilingual and Multicultural Development, 1*(1), 41-47.

Pritchard, R., & Loulidi, R. (1994). Some attitudinal aspects of foreign language learning in Northern Ireland: Focus on gender and religious affiliation. *British Journal of Educational Studies*, *42*(4), 388-402.

Purschel, H. (1981). The evaluation and assessment of oral performances by German English teachers. [Die einschatzung und bewertung mundlicher leistungen durch Deutsche Englischlehrer] *Praxis Des Neusprachlichen Unterrichts, 28*(4), 368-375.

Reese, S. (2008). Teacher education—Growing the future of our profession. *The Language Educator, 3*(6), 22-26.

Ricardo-Osorio, J. (2008). A study of foreign language learning outcomes assessment in U.S.

  undergraduate education. *Foreign Language Annals, 41*(4), 590-610.

Rifkin, B. (1996). Rewarding oral communication in the Slavic languages classroom. *AATSEEL*

  Newsletter, 38(6), 9-11.

Robison, R. (1992). Developing practical speaking tests for the foreign language classroom: A

  small group approach. *Foreign Language Annals, 25*(6), 487-496.

Ross, S. (2007). A comparative task-in-interaction analysis of OPI backsliding. *Journal of*

  Pragmatics, 39(11), 2017-2044.

Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency

  interviews. *Studies in Second Language Acquisition, 14*(2), 159-176.

Rosser, H. (1979). Testing oral communicative skills. *Foreign Language Annals, 12*(5), 371-

  373.

Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language

  class. *Foreign Language Annals, 42*(1), 158-178.

Salaberry, R. (2000). Revising the revised format of the ACTFL oral proficiency interview.

  Language Testing, 17(3), 289-310.

Savignon, S. (1997). *Communicative competence. theory and classroom practice: Texts and*

  contexts in second language learning. (2nd ed.). New York, NY: McGraw-Hill.

Shrum, J., & Glisan, E. 2005). *Teacher's handbook--contextualized language instruction* (3rd

  ed.). Boston, MA: Thomson Heinle.

Silverberg, S., & Samuel, A. (2004). The effect of age of second language acquisition on the

representation and processing of second language words. *Journal of Memory and*

Language, *51*(1), 381-398.

South, C., Rugelbak, J., Usai, P., Kitawaki, N., Irii, H., Rosenberger, J., et al. (1993).

Subjective performance assessment of CCITT's 16 kbit/s speech coding algorithm.

Speech Communication, 12(2), 113-133.

Spolsky, B., Murphy, P., Holm, W., & Ferrel, A. (1972). Three functional tests of oral

proficiency. *TESOL Quarterly, 6*(3), 221-235.

Sprenger, M. (2005). *How to teach so students remember.* Alexandria, VA: Association for

Supervision and Curriculum Development. (ASCD).

Stansfield, C., & Kenyon, D. (1992). The development and validation of a simulated oral

profiency interview. *The Modern Language Journal, 76*(2), 129-141.

Stansfield, C., & Kenyon, D. (1993). Development and validation of the Hausa speaking test

with the ACTFL proficiency guidelines. *Issues in Applied Linguistics, 4*(1), 5-31.

Stansfield, C., Kenyon, D., Paiva, R., Doyle, F., Ulsh, I., & Cowles, M. (1990). The

development and validation of the Portuguese speaking test. *Hispania, 73*(3), 641-651.

Stewart-Strobelt, J., & Chen, H. (2003). Motivations and attitudes affecting high school

students' choice of foreign language. *Adolescence*, *38*(149), 161-170.

Surface, E., & Dierdorff, E. (2003). Reliability and the ACTFL oral proficiency interview:

Reporting indices of interrater consistency and agreement for 19 languages. *Foreign*

Language Annals, 36(4), 507-519.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B.

    Seidhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G.*

    Widdowson. Oxford, UK: Oxford University Press.

Swain, M. (2000). A vygotskian perspective on corrective feedback in L2: The effect of random

    versus negotiated help in the learning of English articles. *Language Awareness*, *9*(1), 34-

    51.

Swain, M. (2001). Examining dialogue: Another approach to content specification and validating

    inferences drawn from test scores. *Language Testing*, *18*(3), 275-302.

Swain, M. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review*

    of Applied Linguistics, 22, 171-185.

Swain, M. (2004). What underlies immersion students' production: The case of avoir besoin de.

    Foreign Language Annals, 37(3), 349-355.

Swain, M. and Lapkin, S. (1995). Problems in output and the cognitive processes they generate:

    A step towards second language learning. *Applied Linguistics, 16*, 371-391.

Szostek, C. (1994). Assessing the effects of cooperative learning in an honors foreign language

    classroom. *Foreign Language Annals, 27*(2), 252-261.

The assessment of standards. (1977). *Modern Languages in Scotland, 13*(May), 100-103.

The National Council of State Supervisors for Languages. *Foreign Language Education for all*

    Students. http://www.ncssfl.org/papers/index.php?allstudents

Thompson, G. (2006). Teacher and student first language and target language use in the

    foreign language classroom: A qualitative and quantitative study of language choice.

    (PhD., The University of Arizona).

Thompson, I. (1995). A study of inter-rater reliability of the ACTFL oral proficiency interview in five European languages: Data from ESL, French, German, Russian, and Spanish. *Foreign Language Annals, 28*(3), 407-422.

Traphagan, T. (1997). Interviews with Japanese FLES students: Descriptive analysis. *Foreign Language Annals, 30*, 98-110.

Troyan, F. (2008). Being authentic: Assessing standards-based tasks in a content-based curriculum. *The Language Educator, 3*(4), 52-54.

Vail, K. (2006). Increased AP test taking raises questions. *The American School Board Journal*, *193*(4), 6.

Van Eunen, K. (2002). "Eine cola und zwei fanta bitte!" The city Germany: Performance assessment in authentic communicative situations. ["Eine cola und zwei fanta bitte!" Sprachstadt Deutsch: Leistungsbewertung in authentischen kommunikationssituationen *Fremdsprachenunterricht, 46*(4), 249-251.

Verkler, K. (1994). Middle school philosophy and second language acquisition theory: Working together for enhanced proficiency. *Foreign Language Annals*, *27*(1),19-42.

Verkler, K. (2008). A study on the motivation of high school Spanish students to continue language study beyond the second year. *Florida Foreign Language Journal*, *5*(1), 19-26.

Wallace, C. (2003*). Critical reading in language education*. New York: Palgrave Macmillan

Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the language of love and stuff': Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal, 28*(4), 503-528.

Yorozuya, R., & Oller, J.,Jr. (1980). Oral proficiency scales: Construct validity and the halo effect. *Language Learning, 30*(1), 135-153.

Young, D. (1985). The relationship between anxiety and foreign language oral proficiency ratings (interview, second language, testing). (Ph.D., The University of Texas at Austin).