

**REPORT ON THE HAWAIIAN ORAL LANGUAGE ASSESSMENT  
(H-OLA) DEVELOPMENT PROJECT<sup>1</sup>**

**ALOHALANI HOUSMAN, KAULANA DAMEG, MĀHEALANI KOBASHIGAWA**  
*University of Hawai‘i at Hilo*

AND

**JAMES DEAN BROWN**  
*University of Hawai‘i at Mānoa*

**INTRODUCTION**

Hawaiian language immersion program (HLIP) education came into existence in the Hawai‘i public school system over 23 years ago. During the first several years of the program, the Hawai‘i State Department of Education (DOE) hired evaluators to conduct extensive qualitative research. Students were assessed in oral language development, reading, writing, and mathematics. It is important to keep in mind that the Hawaiian language immersion movement preceded the charter school movement by about 13 years and therefore was a brand new and foreign concept to the DOE. Although the State Superintendent and the Board of Education (BOE) approved the start of the pilot program in 1987, school officials were somewhat apprehensive of the long-term effect that such a program could have on the ability of students to compete in an English-speaking world. Therefore, the major impetus of the evaluations in the early years of the program was not to understand the acquisition of the Hawaiian language, but rather to inform the state school officials of the general progress of the program to determine whether it should continue from year to year and also to ensure that students were progressing satisfactorily both in Hawaiian and English.

In the 1989-1990 and 1990-1991 school years, the first systematic study to investigate the effectiveness of the HLIP in transmitting the Hawaiian language to a new generation of children was conducted. Even though the study only included the language excerpts of students from one school, the information that Warner (1996) collected during his research is invaluable in that it

---

<sup>1</sup> This project was funded by the Native Hawaiian Education Program of the USDOE, CFDA Number: 84.362A (8/1/2007 to 7/31/2010).

provides the first analysis of grammatical structures produced by HLIP students in the early elementary grades.

In more recent years, especially since the No Child Left Behind Act was written, the DOE has translated or created standardized assessment tests to evaluate students' abilities in the area of reading, writing, and math. However, to date, no formal oral language assessment in the Hawaiian language has been developed by the State office. Since no formal oral language assessment is available, individual HLIP schools and/or teachers have attempted to assess their students informally within the classroom setting. This type of informal assessment aids in evaluating students abilities, which can then be used to inform instruction, but in order to ensure consistency between schools, it is imperative that a statewide standard be established to determine language proficiency levels of students across the board in the HLIP.

The vision statement of the HLIP is as follows: "The Hawaiian language and culture shall be the foundation upon which excellence is built, thereby determining, shaping and guiding the future for all of Hawai'i's people" (Office of Instructional Services, General Education Branch, 1994). One of the major goals of the program is to develop a high level of proficiency in comprehending and communicating in the Hawaiian language. This means that Hawaiian needs to be revitalized to the point where it is a thriving language and can be used in any context of life. If this is the purpose of the program, then it is crucial to assess and evaluate the language development of students on a formal basis. Since no systematic study has taken place in almost 20 years, this study proposes to determine the language proficiency levels of students in the early elementary grades. A comprehensive approach will be taken to assess seven proficiency domains: communicative skill, vocabulary, grammar, pronunciation, fluency, language steadfastness, as well as cultural and linguistic authenticity.

## **BACKGROUND OF THE HAWAIIAN LANGUAGE REVITALIZATION MOVEMENT**

The Hawaiian language once flourished throughout the islands and was deemed important at all levels of Hawaiian society. For centuries, Hawaiian was an oral language, but several years after the missionaries arrived in Hawai'i in 1820, a written language was created. It didn't take long for a well-developed public school system taught through the medium of Hawaiian to be established. By 1831, 50,000 adult Hawaiians were attending over 1,000 native schools. This is

amazing, since it was estimated by Everly (1965) that the population only consisted of approximately 85,000 Native Hawaiians. Teachers of Native Hawaiian lineage were prepared by missionaries to teach in community schools. Public schools, taught through the medium of Hawaiian, produced a very high rate of literacy resulting in Native Hawaiians having the largest volume of written literature of any indigenous people in the world at that time. The literacy rate in the late 1800s surpassed the United States with literacy rates of 84% and 91.2% respectively for full-blooded Hawaiians and part-Hawaiians over the age of six (Hawai‘i General Superintendent of the Census, 1897). Many educators attribute the immense success in literacy growth to the fact that Hawaiian was the respected living language of the home and school during the 19th century.

The prominent status of the Hawaiian language enjoyed by Natives during that time is in stark contrast with the condition of the Hawaiian language over the 100 years that followed. Today, Hawaiian is an endangered language. After the overthrow of the Hawaiian government in 1893, a legislative ban on Hawaiian medium schools followed three years later. Students were severely punished for speaking their native tongue in the school setting. The result was the decline of the Hawaiian language in all facets of life. Hawaiian language use dwindled in the home and the community. According to Dr. Kalena Silva, a 1983 survey estimated that only 1,500 people could speak Hawaiian, most of them elderly (Staton, 2005). Surveys in the late 1980s showed that, after nearly a century of English only schools, the Hawaiian language was nearly extinct and 30% of the Native Hawaiian population were at the lowest level of literacy for all ethnic groups in Hawai‘i (Kana‘iaupuni & Ishibashi, 2003).

In an effort to revitalize the use of the Hawaiian language, a movement began in 1983 with the establishment of the nonprofit ‘Aha Pūnana Leo, Inc. (‘APL) and its first Hawaiian immersion pre-kindergarten program in 1984. When this endeavor began some 25 years ago, only 35 children under the age of 18 were fluent in Hawaiian (‘Aha Pūnana Leo, Inc., no date). It is through the Pūnana Leo experience that young families, who had the same desire of raising their children speaking Hawaiian, came together in a shared vision, “E ola ka ‘ōlelo Hawai‘i,” which means, “the Hawaiian language shall live” (‘Aha Pūnana Leo, Inc). Strong relationships developed between the families through participation in school service, parent meetings, weekly language classes, and fundraisers. The Pūnana Leo is what tied stakeholders together and kept the Hawaiian language movement progressing to the next level. With their children soon to graduate

from the preschool, it became apparent that parents wanted a means for them to continue their education in Hawaiian at the elementary school level. Parents understood that without language, the culture will eventually cease to exist, and the native people known as Hawaiians would eventually be no more.

The ‘Aha Pūnana Leo Board and parents presented testimonies before the Board of Education and the State Legislature. In 1986, the law was changed to allow the Hawaiian language to be the medium of instruction in the public school system once more. Through the united efforts of Native Hawaiian voices, the Hawaiian language immersion pilot program was approved to begin in Fall 1987 (Kawai‘ae‘a, Housman, & Alencastre, 2007). Superintendent Toguchi agreed to open two K-1 combination classes, one on Hawai‘i island and one on O‘ahu, with the intention of adding a new grade each year. This was a monumental step in the language revitalization movement. Since the program was only approved as a pilot program, the success of the students would determine the longevity of Hawaiian language immersion in the public school system.

As the program overcame the first hurdle of approval, other obstacles quickly became visible. Where would the curriculum materials and books come from? Which principals would be willing to allow a Hawaiian language immersion program in their school? Who had a teaching certificate and could speak Hawaiian? How would the program be funded? The odds were stacked against the start of the program, and yet within weeks everything fell into place.

Each year, over a five-year period, the ‘Aha Pūnana Leo and parents went back to the Board of Education to request approval for additional grades and supportive program policies. The annual continuation of the pilot project was contingent on the academic success of the program as determined by external evaluators who came into the classroom to observe the class, write notes, take pictures, assess individual students, and then write a qualitative report for the State Superintendent and Board of Education to review. A milestone was achieved on June 15, 1989, when the pilot status was changed to a permanent program status. Even then, however, the program was only approved to continue until the 6th grade. It wasn't until 1992 that the Board of Education approved the extension of the program to grade 12 (Office of Instructional Services, General Education Branch, 1994).

Through the relentless effort and commitment of Hawaiian leaders, teachers, and families, the HLIP has continued to grow and expand over the past 23 years. The program has grown from

two K-1 classes with 34 students in 1987 to the current 20 schools with approximately 2,000 students statewide (Hale Kuamo‘o-University of Hawai‘i at Hilo, 2010).

### **REVIEW OF RELEVANT LITERATURE**

The Hale Kuamo‘o Hawaiian Language Center located at the University of Hawai‘i at Hilo received a Native Hawaiian Education grant from the federal government in 2007. The central goal of the grant project, entitled ‘Ōlelo Ola (the Living Language), is to develop a high level of Hawaiian language oral proficiency among Hawaiian language immersion students enabling them to meet and exceed standards in Hawaiian. There are five main objectives of the grant (Ka Haka ‘Ula O Ke‘elikōlani, 2007, p. 6):

1. To investigate current research on oral language development and proficiency that benefits Hawaiian language immersion students in grades K-3.
2. To develop assessment resources and curricula that support Hawaiian oral language development and proficiency for K-3 HLIP students.
3. To conduct ongoing assessment of the Hawaiian language oral proficiency of K-3 HLIP students by expanding and adapting research.
4. To strengthen the oral language proficiency of K-3 HLIP students by providing teacher preservice and inservice training.
5. To support families of K-3 HLIP students, and other interested community members, in developing Hawaiian language oral proficiency and literacy skills within the home environment.

This report in its entirety will only focus on the first three objectives of the grant. The research team also completed the last two objectives, however, they will not be covered in this document. The first objective will be addressed in the Review of Literature. The second objective will be explained in the Methodology section. The third objective will be described in the Methodology and Results sections.

The first task of the research grant team was to investigate current research on oral language development and proficiency. Following are discussions of the five assessments that provided major contributions to the construction of the Hawaiian Oral Language Assessment (H-OLA) and the rubric that was used to score the language proficiency level of HLIP students. These

assessments were chosen based on the similarities of language revitalization goals found amongst indigenous people as well as their relevance to an immersion educational context.

### *Hawaiian Language Immersion Program Qualitative Assessment*

A qualitative assessment done in the first year of the HLIP program (1987-1988) consisted of classroom observations and student interviews that focused on the areas of oral language ability, writing skills, reading skills, oral reading skills, and listening comprehension. The objective was to measure how proficient the students were in the Hawaiian language as well as in the English language. Two formal assessment tests were used: The Language Proficiency Measure (LPM) and the Peabody Picture Vocabulary Test (PPVT) (Slaughter & Watson-Gegeo, 1988).

The LPM was a qualitative assessment used to assess the oral language proficiency of 24 out of 34 total students, or 70% of the student population (Slaughter & Watson-Gegeo, 1988). This test used “conversation and storytelling from a wordless storybook to elicit student discourse in their first and second language” (Slaughter & Watson-Gegeo, 1988, p. 11). In order for the students to feel more comfortable during testing, they were assessed in pairs. This test was conducted in both English and Hawaiian, however, by different raters and not on the same day (Slaughter & Watson-Gegeo, 1988).

The PPVT is a “nationally normed test of receptive oral language vocabulary, a ‘listening’ or ‘hearing’ vocabulary, that requires students to point to the correct noun, adjective, or action verb (gerund) out of a choice of four pictures” (Slaughter & Watson-Gegeo, 1988, p. 12). A formal version of the test was not translated into the Hawaiian language for the following reasons:

- a) there are no norms for the test, or any other children’s vocabulary test in Hawaiian, and therefore any results could be suspect and misleading, b) many terms on the test would require more than one word in a valid Hawaiian translation, and the PPVT is a test that tests vocabulary out of the context of real discourse, c) it is unknown whether or not the words on the test are as useful and familiar in Hawaiian as they are in English, and d) some of the words on the PPVT do not have a straightforward translation into Hawaiian and the authors of the PPVT abjure against removing any words for any reason from the test (Slaughter & Watson-Gegeo, 1988, p. 13).

However, due to its ease of use and its potential for providing valuable information, an informal Hawaiian language version of this test was conducted in which the examiner asked the students

to produce vocabulary through speaking rather than just listening and pointing to objects, which was considered a much harder task (Slaughter & Watson-Gegeo, 1988).

The results of the LPM showed that 22/24 students were considered moderate to very proficient in the Hawaiian language. The other two students were considered *functional*, as they had interjected English 50% of the time during the conversation segment, and 15% of the time during the narrative segment of the exam. All 24 students were able to converse well in the English language, having no problem with listening comprehension or speaking fluency (Slaughter & Watson-Gegeo, 1988).

The results of the PPVT showed that the majority of the students who were tested entered the HLIP with a below-average English vocabulary for their age group. However, it is interesting to note that a post-test conducted later in the year showed that the students' English vocabulary had improved, and the students' scores were close to being on a par with the national average for students in grades K-1 (Slaughter & Watson-Gegeo, 1988).

The oral language assessment done in the second year of the program (1988-1989) focused more on the fluency and proficiency of the HLIP students in the Hawaiian language rather than on their fluency and proficiency in the English language. The assessment tool used was the Student Communicative Competencies Inventory, which required the teachers of grades K-2 to observe the students' language use in the classroom, and then rate their oral language ability on the following scale: Fluent/Proficient, Moderately Fluent/Proficient, Limited Fluency (Slaughter, Warner, & Palmeira, 1989). The results of the second year assessment showed that “approximately all but a few students were evaluated as fluent, or moderately fluent, and that approximately 70-80% of the students were rated as fluent/proficient in Hawaiian” (Slaughter et al., 1989, p. 18). The data indicated that those who were described as having “limited fluency” were students in grades K-1, suggesting that by the end of grade 2, most HLIP students had reached a high level of fluency in Hawaiian. The data also showed that students had a higher level of fluency in small group settings and demonstrated more proficiency in informal conversations than in directed speech, i.e., storytelling, retelling events, explanations, and giving elaborated responses to teacher questions (Slaughter et al., 1989).

In the third year of the program (1989-1990), a more in-depth assessment was conducted to determine the basic language use of the students. Interviews of students in grades 2-3 were recorded and transcribed and then used to determine the students' proficiency in 30 basic

grammatical/linguistic structures and features. Out of 14 students assessed in grade 3, 12 students (86%) were rated as moderately to highly proficient and only 2 students (14%) were considered to have a moderately low proficiency. In grade 2, 11 out of 16 students (69%) were rated as moderately to highly proficient and the remaining 5 students (31%) were rated as having a moderately low to basic communicative proficiency. (Slaughter et al., 1989)

A similar assessment was done in the fourth year of the program (1990-1991). Students in grades 2-4 were rated on a three-level scale: proficient, moderately proficient, or functionally proficient. Although the evaluation team intended to analyze 44 total grammatical features, it was discovered that not all the interviews produced enough data on all 44 features. Therefore, in grade 4, only 35 grammatical features were analyzed, and the students demonstrated high levels of proficiency in 26 out of 35 (74.3%), and were deemed as developing proficiency in nine out of 35 (25.7%). In grade 3, thirty-four features were analyzed, and the students demonstrated high levels of proficiency in 26 out of 34 (76.5%), and were rated as developing proficiency in eight out of 34 (23.5%). Students in grade 2 demonstrated high proficiency in 29 out of 38 total features analyzed (76.3%), and demonstrated as developing proficiency in nine out of 38 (23.7%). (Slaughter, Lai, Warner, & Palmeira, 1990)

One factor that posed a significant challenge for the HLIP in its first few years of operation was the limited availability of quality curriculum materials in the Hawaiian language. During the program's first year, it was observed at one school that the teachers would translate borrowed library books into Hawaiian, paste the translation over the English for use in the classroom, and then remove the translation when the books were returned to the library (Slaughter & Watson-Gegeo, 1988). Indeed, it was a common practice for teachers and parents to use their own money to purchase English materials (story books, math textbooks, geography textbooks), translate them into Hawaiian, type up the translation, and then paste the typed Hawaiian translation over the English for use in the classroom.

Anecdotes and observations such as these prompted the evaluation teams during the years of 1987 to 1991, to frequently recommend that more funds be provided to produce quality curriculum materials in the Hawaiian language. In 1989, the Hawai'i State Legislature established the Hale Kuamo'o Hawaiian Language Center, which would serve as the main curriculum development center for HLIP schools. It is important to note, however, that since the State of Hawai'i did not provide Hale Kuamo'o with sufficient funds to adequately staff the



Hawaiian Language Center, the Center has largely relied on grants to provide books, curriculum resource materials, assessments, and teacher in-service training workshops over the past 21 years.

***I Ola Ka ‘Ōlelo I Nā Keiki: Ka ‘Apo ‘Ia ‘Ana O Ka ‘Ōlelo Hawai‘i I Nā Keiki Ma Ke Kula Kaiapuni***

The first systematic study to investigate the effectiveness of the Hawaiian language immersion program (HLIP) in transmitting the Hawaiian language to a new generation of children was conducted during the 1989-1990 and 1990-1991 school years. The study was longitudinal and examined oral speech data of students in Kindergarten through grade 4 at one school over a two-year period. According to Warner (1996), the data, primarily collected from 30 to 40 minute oral semi-structured interviews, was audio-taped, transcribed, and analyzed using both quantitative and qualitative methods, and compared with similar data from adult native speakers.

The majority of HLIP students are second language learners of Hawaiian. In addition, children in the program primarily learn from second language learners and speakers, namely their teachers and parents. The context in which children learn Hawaiian prompted Warner (1996, p. 41) to propose the following research questions:

1. How well are particular fundamental structures of Hawaiian grammar, selected as a focus for this study, used in spoken discourse by Hawaiian immersion children after one to three years of instruction?
2. How do the Hawaiian immersion children compare with adult native speakers of Hawaiian with respect to the particular grammatical structures selected for study?

In Warner's (1996) study, Hawaiian language immersion students were compared to adult native speakers of Hawaiian due to the lack of native speaking children outside of the Ni‘ihau community. Access was not possible on Ni‘ihau and limited on Kaua‘i.

The learning of Hawaiian includes several domains such as phonology, pronunciation, vocabulary development, along with sociolinguistic and cultural knowledge. Although all of the aforementioned domains of language are important, only a limited number of features can be examined in any one study. Therefore, the researcher of the study chose to focus on the basic

syntax of the language. The selection of particular syntactic structures of Hawaiian was based on the following criteria:

The syntactic feature should perform an important, if not primary function in the language (and is not rare or highly formal); and that it presents contrast in some major respect within the Hawaiian immersion children's native language, Hawai'i Creole English or another variety of North American English, and the target language, Hawaiian. (Warner, 1996, p. 41)

The Hawaiian syntactic structures that were investigated in the study were divided into three major areas. The first area is related to word order in Hawaiian. The second area deals with noun phrases and noun phrase markers. The third area has to do with verb and predicate head related phenomena (Warner, 1996).

The results of this study were very specific, detailed, and comprehensive. The researcher presented findings and a discussion in three separate chapters that addressed the major areas of study that were mentioned above. Thirty-five tables of data were introduced and expounded upon within the three chapters of findings with an additional twenty-eight tables displayed in the appendices. Following is a succinct summary of the major finding of this study, which compared the conventional use of Hawaiian language syntactic structures by HLIP children and adult native speakers:

Findings indicated that the children were near the levels of conventional use found for the native speaker group for the majority of the fundamental aspects of Hawaiian syntax examined, and were above the group-acquisition criterion of 90%. Furthermore, their nonconventional uses were largely systematic in nature, reflecting the use of prior knowledge (experience including L2, but dominated by L1) as a major strategy employed in learning Hawaiian (related to substrate L1 transfer). In this process, children sometimes created new structures, and regularize perceived irregularities (aspects) of Hawaiian via the overgeneralization of certain rules, often surfacing as Hawaiian (near-) equivalents of English expression. (Warner, 1996, p. xii)

Warner's findings are very interesting in the sense that although Hawaiian language immersion children were able to speak Hawaiian near the levels of conventional use found for adult native speakers, the thinking behind the construction of their Hawaiian grammatical sentences sometimes resembled the structure of English, which is the first and dominate language for most Hawaiian language immersion students. This is an important finding, since a major goal of the

immersion program is not only to cultivate students who can speak fluently in Hawaiian, but also to foster the ability to construct Hawaiian grammatical structures using Hawaiian thought and perspectives.

### ***Kaiaka Reo***

The New Zealand Ministry of Education commissioned the University of Waikato in 1999-2001 to develop a Māori language proficiency assessment tool in the form of proficiency tests for Year Five (8-10 year old) and Year Eight (11-13 year old) Māori immersion students (Edmonds, 2008). The project became known as *Kaiaka Reo*. The underlying ethos of the project is the belief that knowledge is embedded in language and culture. The project strived for excellence by developing a research instrument and a process for Māori medium education by Māori, for Māori, and in the Māori language. According to Edmonds (2008), the proficiency tests were developed by a concerned group of Māori educators who “set out to honor the language of their ancestors and support the young learners learning through the medium of Māori, in cooperation with the New Zealand Ministry of Education, *Te Pua Wānanga ki te Ao* (The School of Māori and Pacific Development) of the University of Waikato, and the Māori medium education sector” (p. 32). The tests comprised a battery of six test sets for Year Five and one test set for Year Eight. Each test set had four components: listening, speaking, reading, and writing. The main purpose of the tests was to reflect the underlying Māori language competence of the students by their performance on the assessment tests. Edmonds (2008) explains that “the test items were to take cognizance of theory related to second language testing and methodology, grammatical competence, discourse competence, phonological competence, and Māori culture” (p. 35).

The following is a summary of the background information that classroom teachers provided regarding the Year 5 Māori immersion students who participated in the Kaiaka Reo Project (Littler, 2001, p. 3):

- The 277 Year 5 students tested were mostly in the 8.5 to 10.5 year age range, with almost equal numbers of girls (49%) and boys (51%).
- Almost all students had attended Kōhanga Reo (Māori immersion preschools), with 50% having 3 or more years of Māori language
- 55% of the students had spent four or more years learning in Māori

In addition, the following is a summary of the background information that classroom teachers provided regarding the Year 8 Māori immersion students who participated in the Kaiaka Reo Project (Littler, 2001, p. 4):

- The 202 Year 8 students tested were mostly in the 11-13 year age range, with almost equal numbers of girls (49.5%) and boys. (50.5%).
- Almost all students had attended Kōhanga Reo, with 62% having 3 or more years of Māori language.
- 81% of the students had spent 4 or more years learning in Māori.

In a report on Kaiaka Reo data (Littler, 2001), statistics showed that proficiency levels of Year 5 students were higher in the speaking (56.3 Mean) and listening (56.2 Mean) categories as compared to writing (49.3 Mean) and reading (40.0 Mean). Similar results were found with the Year 8 students. Statistics showed that proficiency levels of Year 8 students were also higher in the speaking (63.6 Mean) and listening (52.0 Mean) categories as compared to writing (47.3 Mean) and reading (51.3 Mean).

In addition to statistics cited in the Kaiaka Reo data report regarding proficiency levels of students in speaking, listening, writing, and reading, additional findings were expounded on in the report. Following are background factors that may have affected performance (Littler, 2001):

- Influence of gender - At both Years 5 and 8, girls are performing better than the boys in the overall results, with the effect statistically significant in the writing test at Year 5 and especially in the writing and reading tests at Year 8. (p. 18-19)
- Influence of classroom use of Māori language - Classroom use of Māori language correlates with an increased performance across all tests. On average, Littler (p. 23) shows that Year 5 students who do not use or “sometimes” use Māori in the classroom are performing at 5% less than the peers who use Māori most of the time or all the time, with the latter group averaging about 50% over the four skills. Littler (p. 24) shows that Year 8 students who do not use Māori, average 10% less than those who use Māori “sometimes”. Those who use Māori “sometimes” average 10% less than their peers who use Māori most of the time or all the time. Those students who use Māori most or all the time perform at the 60% level.
- Influence of home use of Māori language - For those students at Year 5 for whom Māori is used in the home, average proficiency levels were 10% greater than those who did not

have Māori in the home (p. 26). For students at Year 8, the effect of home use on proficiency levels vanishes (p. 27).

### ***Cherokee Immersion Language Assessment (C-PILA/C-KILA)***

Cherokee belongs to the Iroquoian language family, is polysynthetic, and has a syllabary writing system of 85 characters that was developed by Sequoyah, a Cherokee silversmith, in 1821. There are an estimated 9,000 speakers of Cherokee in Oklahoma (Linn, 2004). However, a recent survey of registered tribal members revealed that only 11% of the respondents consider themselves to be fluent in the language and most of those speakers are over the age 40 (Cherokee Nation, 2003). More importantly, the survey revealed that Cherokee is no longer taught as the mother tongue to children in the home (Peter & Hirata-Edds, 2006). Based on UNESCO's Language Vitality Scale of Intergenerational Language Transmission, unless something drastic is done to reverse the language shift, in the next three decades, only a few fluent speakers of Cherokee will remain (Peter & Hirata-Edds, 2006).

In an effort to save the language, the first Cherokee language immersion preschool opened its doors to 17 three and four-year-olds in 2001 in a small town called Tahlequah in the state of Oklahoma (Peter & Hirata-Edds, 2006). In 2005, an immersion kindergarten class was added. First grade immersion started in 2006 and second grade in 2007 (Raymond, 2008). Today, the school continues to educate students in their native Cherokee tongue from preschool to grade 5 (Cherokee Nation, 2009). According to Peter, Sly, and Hirata-Edds (2008), the ultimate goal of the Cherokee Language Immersion Mission is “for children to acquire the Cherokee language in such a way that it will become an integral part of their lives and their knowledge about the world around them” (p. 3). Following are the Four Targeted Immersion Areas of the Cherokee Immersion School (Peter et al., 2008, p. 3):

- Cherokee Way: Language in a broader sense, including culture, community, and spirituality
- Curriculum: Guiding framework; instructional goal setting; incorporation of standards for language, culture, and academic content
- Instruction: Teaching and learning infant through adult; professional development for those directly and indirectly involved with instruction

- Assessment and Evaluation: Documenting growth and development of learners' language/academic skills; overall program evaluation

In order to evaluate the effectiveness of the program, several formal language assessment instruments were developed by Lizette Peter and Racy E. Hirata-Edds (Peter & Hirata-Edds, 2006) who worked closely with classroom teachers and language staff from the Cherokee Nation Cultural Resource Center. The C-PILA: Cherokee Preschool Immersion Language Assessment is designed for children between the ages of two and five who are enrolled in a Cherokee language immersion classroom. The purpose of the assessment is to measure the extent to which children learning Cherokee through immersion have developed skills to communicate competently—including knowledge of vocabulary, ability to comprehend questions and commands, and ability to respond appropriately to questions and commands either verbally or through action (Peter & Hirata-Edds, 2006). Feedback from the results of the assessment allowed teachers to recognize that children needed more opportunities to use the language in meaningful ways (Peter et al., 2008). In addition to the C-PILA assessment, the C-KILA: Cherokee Kindergarten Immersion Language Assessment was also developed. The C-KILA was designed for Kindergarten students of the Cherokee language immersion classroom. The question that guided the process is, “What should children be able to do in the language by the end of Kindergarten and after two or three years in immersion?” (Peter et al., 2008, p. 15). The C-KILA, with revisions, was also used with the first and second grade students.

Formal language assessments are tools that help determine the communicative competence of immersion students. Analysis of the information is an excellent way to inform instruction, as was done in the case of the Cherokee language immersion school. Conducting several professional development workshops after the initial pre-test was an effective means to help teachers, and in the end students, make significant improvements in the acquisition of language as was demonstrated in the post-test evaluation (Peter & Hirata-Edds, 2006). Assessments used for this purpose have a positive backwash on the program as opposed to high stakes testing which typically have punitive consequences on schools. Educators of the Cherokee Language Immersion School along with linguistic and educational professors at Kansas University are to be applauded for the collaboration and insights that provided teachers with information they needed to actualize the goals of their instruction and to make enlightened educational decisions that successfully impacted the development of competent language skills of their students.

### ***Student Oral Proficiency Assessment (SOPA)***

The Center of Applied Linguistics developed the Student Oral Proficiency Assessment (SOPA). The SOPA, which was developed in 1991 for children in Grades 1-4 in a Spanish partial immersion program, is interactive and assesses students in pairs. In 1996, the interview format and rating scale of the SOPA were adapted for non-immersion Foreign Language in the Elementary School (FLES) programs. Over the years, the SOPA has been used with students in Grades K-8 in an increasing number of language programs such as FLES, Foreign Language Exploratory or Experience (FLEX), and immersion (Thompson, 2006). According to Thompson (2006), the purpose of the assessment is to determine the highest proficiency levels in speaking and listening comprehension that students can sustain at a particular point in time.

The goal of the SOPA assessment is to show what the students can do in the target language rather than focus on what they cannot do. An interview format is used. The interview consists of a series of games or tasks with various levels of difficulty that elicit both academic and social language. Interviews take approximately 15-20 minutes to complete and are conducted entirely in the target language. There are two SOPA scripts. One script was developed for non-immersion programs such as FLES, the other script was developed for immersion programs. Following is an outline of the Immersion SOPA (Thompson, 2006, p. 3):

#### Task 1: Fruits & Other Foods

Goal: TPR with fruits and other foods - Put students at ease with simple listening comprehension tasks first, followed by speaking.

#### Task 2: All About You

Goal: Answer questions - Give students opportunities to create with language related to familiar topics.

#### Task 3: The Life of a Plant

Goal: Describe - Give students opportunities to use academic terminology and to create at the sentence level.

#### Task 4: Story Retelling

Goal: Describe, narrate - Give students opportunities to speak in past time at the paragraph level.

#### Task 5: School Rules

Goal: Support/Defend an opinion, hypothesize - Give students opportunities to speak at the advanced level, using formal discourse extended beyond the paragraph level.

Wind-Down: TPR or Easy Questions

Goal: End at students' comfort level.

The SOPA has a choice of two rubrics that can be used to score the interview. Thompson (2006) explains that the COPE/SOPA-Rating Scale (RS) consists of nine proficiency levels and the SOPA Rating Scale (SOPA-RS) consists of six proficiency levels. Each proficiency level of the COPE/SOPA-RS has four skill areas: oral fluency, grammar (speaking), vocabulary (speaking), and listening comprehension. The six-level scale is a subset of the nine-level scale, but it has only two skill areas: oral fluency and listening comprehension. Typically, students in immersion programs have more advanced skills, therefore, the COPE/SOPA-RS is used to rate students. Evaluators should keep in mind that the SOPA results are only one indicator and should be used in conjunction with teacher observations and other evaluations of student work.

## **PURPOSE OF PROJECT**

A central objective of the 'Ōlelo Ola project was to develop a detailed and comprehensive oral language proficiency assessment to collect baseline data on the oral language proficiency levels of HLIP students in grades 1-3 at seven participating schools. This was to be accomplished through the creation and utilization of a standards-based assessment tool and an oral language proficiency rubric.

It is important to note that the oral proficiency level of teachers in the classroom is directly related to the language development and proficiency of the students. This is highlighted in the HLIP Program Guide (Office of Instructional Services, General Education Branch, 1994). It emphasizes the importance of teacher oral proficiency and also the connection between oral proficiency and literacy by stating the following:

A skilled, highly proficient teacher can provide the best model of appropriate Hawaiian language usage as well as design curriculum that optimizes language learning for the new speaker of Hawaiian. This enhances students' skills in all aspects of language use—speaking, reading, and writing—as well as minimizes the need to spend time correcting inappropriate use of Hawaiian in subsequent years. (p. 7)



Therefore, of central importance to the ‘Ōlelo Ola project is the ability of the assessment instruments developed to provide continuous feedback for improvement in teaching and learning, as well as a summative evaluation to provide a quantitative measure of systemic program growth.

Since its inception in 1987, the HLIP has grown significantly and the issue of oral language proficiency has not been sufficiently addressed. Thus the ‘Ōlelo Ola team posed the following three research questions:

1. What are the most important aspects of oral language development that should be assessed?
2. What is the Hawaiian oral language proficiency level of Hawaiian Language Immersion Program (HLIP) students in grades 1-3?
3. Once baseline data is collected, what can be done to improve the Hawaiian language proficiency level of students?

The ‘Ōlelo Ola team chose to focus on the oral language proficiency of grades 1-3 not only because of the stipulations of the federal grant with which it was funded, but also to be able to assess and intervene to improve the students’ proficiency level at a critical stage of oral language development. It was decided that kindergarten level students would not be assessed due to the length of the test and their developmental limitations. Also, in order to complete the project in accordance with the timetable outlined in the grant, the project would assess a broad representative sample of students from various schools on the different islands rather than attempting to assess all prospective study participants. The sample selected would be drawn from the three types of schools in the HLIP: charter, laboratory, and DOE public schools.

## METHODOLOGY

This section is divided into four major categories: (a) participants, (b) procedures of test development, (c) materials, (d) test administration steps, and (e) rating steps.

### *Participants*

The participants in this project were 270 students from seven HLIP schools located on four different islands. They ranged in age from 7 to 10 with a mean of 7.93 years and standard deviation of .85 years. In terms of gender, 123 were male and 147 were female. They were in

grades 1 ( $n = 110$ ), 2 ( $n = 86$ ), and 3 ( $n = 74$ ). Seventy attended laboratory schools, 100 were in charter schools, and 100 were in other DOE immersion schools. The seven schools were coded anonymously here 1 to 7, but the distribution of students in schools was as follows: 1 ( $n = 75$ ), 2 ( $n = 45$ ), 3 ( $n = 25$ ), 4 ( $n = 8$ ), 5 ( $n = 17$ ), 6 ( $n = 53$ ), and 7 ( $n = 47$ ). A total of 109 students had previous Pūnana Leo (Hawaiian immersion pre-school) experience, while 161 did not. The number of years in Hawaiian language schools were as follows: 1 year = 8, 2 years = 74, 3 years = 66, 4 years = 83, 5 years = 27, 6 years = 8, and 7 years = 3 (data in this category was not provided for one student). The amount of Hawaiian language use in the home was rated on a 1-6 scale (where 1 = low and 6 = high), but there was one zero given, so there were seven different ratings overall: 0 ( $n = 1$ ), 1 ( $n = 63$ ), 2 ( $n = 77$ ), 3 ( $n = 77$ ), 4 ( $n = 35$ ), 5 ( $n = 11$ ), and 6 ( $n = 6$ ). The language levels of the students were rated by their teachers on a 1-3 scale (where 1 = low and 3 = high), but there were two zeros given, so there were four ratings overall: 0 ( $n = 2$ ), 1 ( $n = 60$ ), 2 ( $n = 154$ ), and 3 ( $n = 54$ ).

There were five raters in this study. Two are professors of Ka Haka ‘Ula o Ke‘elekōlani College of Hawaiian Language and three are members of Hale Kuamo‘o’s ‘Ōlelo Ola grant team. The raters were the same people who wrote, developed, and revised the tests investigated in this project. In their roles as raters, each administered the test to between 47 and 65 students and scored the same students for both the Extended Response Items and the Short Oral Response Items. These raters were coded as: rater 1 ( $n = 47$ ), rater 2 ( $n = 52$ ), rater 3 ( $n = 65$ ), rater 4 ( $n = 47$ ), and rater 5 ( $n = 59$ ).

### ***Procedures of Test Development***

The ‘Ōlelo Ola team was made up of individuals that specialize in Hawaiian language immersion education, and/or Hawaiian language and culture. In order to develop an effective assessment tool, it was essential that an advisory committee consisting of experts on language assessment, linguistics, and Hawaiian language and culture be invited to participate with the ‘Ōlelo Ola team in the development of an assessment tool appropriate for the target grade levels from its draft stages until its finalized version.

An important part of the process in developing the Hawaiian Oral Language Assessment (HOLA) tool along with an appropriate scoring rubric was looking at other oral language assessment instruments that were developed by other experts. The ‘Ōlelo Ola team benefited

from the fact that Katarina Edmonds was completing her doctoral thesis in Hilo during the same time that the Hawaiian assessment was being developed. The team gleaned insights from her knowledge, experience, and expertise regarding Māori immersion, linguistics, and assessment. Katarina played a significant role as the project manager in the development of the Kaiaka Reo Māori Language Assessment. The oral language section of the Kaiaka Reo laid the foundation for the development of the story-telling picture series (described in the Materials section) of the Hawaiian Oral Language Assessment (H-OLA). In addition, the Kaiaka Reo analytical scale was analyzed in the process of developing a scoring rubric for the Hawaiian assessment.

The H-OLA development team felt it was also necessary to include a short response section of the test, which looked at specific components of Hawaiian grammar. Therefore a close examination of the Cherokee Kindergarten Immersion Language Assessment (C-KILA) was undertaken. The Cherokee assessment inspired the development of items included in the short response section of the Hawaiian assessment such as identifying objects (nouns), identifying actions (verb, verb marker, pronoun), using locatives, and following commands.

Lastly, the development team investigated the Student Oral Proficiency Assessment (SOPA) developed by the Center of Applied Linguistics (CAL). Several team members were familiar with the SOPA due to a class offered to Hawaiian language professors and immersion teachers in Spring 2007. During the semester class, Dee Tedick and Tara Fortune from the Center for Advanced Research on Language Acquisition (CARLA) elucidated several assessments that could be used to assess oral language proficiency levels of immersion students. Based on the age and task appropriateness of the SOPA, a similar task construct was created for the H-OLA. The finalized version of the H-OLA consists of four parts that comprise open-ended (Parts 1 & 4) and form-focused (Parts 2 & 3) oral prompts.

1. Introduction/Interview
2. Listening Comprehension & Short Response (identifying objects, demonstrative pronouns, and location)
3. Listening Comprehension & Short Response (verbs, verb markers, personal pronouns)
4. Story-telling picture series (2 series)

***Test piloting.*** Piloting was done in March 2009 at two Hawaiian language immersion program sites, including one laboratory school and one charter school. Teachers of grades 1-3 from the participating schools were asked to select three students from their class to participate

in the pilot study, one from each level of oral language proficiency (low, average, and high). There were 21 students that participated: nine laboratory school students and 12 charter school students. All five evaluators participated in the pilot testing, individually assessing at least one student from each grade level and each level of proficiency.

**Test Revisions.** During pilot testing, three picture series were used: the lei series, the slipper series, and the beach series. This was done to determine which two of the three picture series would elicit the most language from students and would produce the most similar results. It was decided that for the formal test, the lei series and the slipper series would be used. Each series was used to elicit extended oral responses from approximately half of the students during the first assessment with all seven schools in spring 2009 with the understanding that during the follow-up assessment one year later, each student would be evaluated using a different series from the prior year to ensure that students would not memorize the pictures.

One section of the pilot test, oral story retelling based on a wordless book, was removed from the assessment. It was apparent that not all students were familiar with the story chosen, so instead of retelling the story, students had to create a storyline, and those familiar with the story had an unfair advantage. Furthermore, pilot testers observed that the wordless book prompt elicited basically the same type of response as the picture series storytelling prompt. However, during the formal testing, when time permitted, students were asked to tell the story of the wordless book using the best language possible. This information was not scored, but was used in an error analysis report that included the grammatical strengths and weaknesses found in testing. An individual report for each school was given to participating schools and teachers.

In the pilot study, testing was done over two consecutive days in two sessions (approximately 20 minutes per session). Due to time constraints, some revisions were made to shorten the length of the test, keeping the most valuable, reliable or challenging items, and removing duplicate items. To aid in the selection of items to be removed, the form-focused sections of the test were scored and data analysis was done to determine items that were unreliable or not essential to the outcome of the test. Some items were too easy, and therefore unnecessary, and some items were too difficult, and deemed unreliable for the assessment, and some duplicate items were evident. As a result of this analysis, a section used to build confidence was shortened from 14 to 5 items; a section identifying objects, demonstrative pronouns, and locations was reduced from 30 to 18 items; and a section testing verbs, verb markers, and pronouns were removed, reducing that

section from 48 to 27 items. Using the revised test, the administrators were able to complete each assessment in a single 15-30 minute session.

### ***Materials***

The Hawaiian Oral Language Assessment (H-OLA) includes materials for long response and short response sections of the test. The long response sections were developed to elicit open-ended responses that would demonstrate the language being used by students in the target grades. The short response sections were developed to test the students' knowledge of specific sentence structures and part of speech categories.

The long response section of the test includes two different story-telling picture series tests. Each picture series consists of a set of six laminated, color-illustrated picture cards produced by a professional graphic artist following the conceptual guidance of the test development team. Used as prompts for open-ended oral storytelling, the two sets of cards are referred to by the development team as the lei series and the slipper series. In addition, a set of 3 color-illustrated laminated picture cards, referred to as the birthday series, was used for practice with the students in preparation for the official independent story-telling assessment.

The short response section of the test includes an array of manipulatives, selected to assess familiarity with nouns, and a set of laminated, color-illustrated action pictures, which are organized as a spiral bound flip chart.

An instruction booklet was developed by the team to ensure consistency during the administration of the test. Each booklet included detailed instructions on test administration, a script for evaluators, a list of needed supplies, and established goals for each task.

Each assessment started with introductions between the test administrator and the student. The test administrator introduced him/herself (name, names of immediate family members including sibling relationships foreign to the English language, age, place of birth and residence, pets, and favorite activity). The student was then asked to introduce him or herself in the same manner without prompts. If the student could not complete the entire task independently, the administrator would assist the student with scripted prompts. The test administrator then asked follow-up questions to check for understanding of Hawaiian sibling relationship terminology. Finally, an attempt was made to engage the student in casual conversation with the goal of

eliciting as much language as possible while putting the student at ease and establishing comfort and rapport with the administrator.

The next unscored section of the assessment was a simple task that required students to point to familiar objects on the table. This was included to help students feel comfortable and confident about their ability to complete the assessment.

In the first part of the short oral response section of the test, students were asked to first identify an object using the correct noun and demonstrative pronoun and then identify its location in relation to a plastic box. The objects used for this section (spoon, shark, lei, etc.) were selected for their likely familiarity to students in these grade levels. Nine practice items, which were not scored, preceded the 18 formal items of this portion, including six items each for nouns, demonstrative pronouns, and locatives. Practice items were used to ensure that students understood what was expected of them in the task in order to get a true assessment of students' language abilities.

In the second part of the short oral response section, students were shown successive color illustrations and asked to state what action was being done (using the appropriate Hawaiian pronoun) in each picture. Following the set of pictures, the student and/or administrator performed simple actions (waving, clapping, building a house from blocks). For each action, the student was asked to state the action being done and by whom (using pronouns). Nine practice items preceded the 18 formal items, including six items each for verbs, verb markers, and pronouns.

The extended oral response sections of the test included the introductions described in the first task as well as a picture series prompt that was used in the last task of the formal assessment. For the picture series prompt, students were asked to tell a story about six picture cards arranged in a consecutive series, using the best language possible. From the two picture series that were determined following the pilot testing, one was randomly assigned to each student. A three-picture series was used as practice for this section of the test. Administrators provided coaching as needed, only during the practice series. Students were also asked to predict what would happen after the last picture for both the practice series and the formal assessment prompt that followed the practice item.

### ***Test Administration Steps***

Students were randomly assigned to an administrator fluent in the target language and tests were conducted individually, with one administrator and one student. It was decided that students would be tested individually so that each student could interact with a test administrator without being interrupted or dominated by another student, or feel uneasy about responding to questions in front of another student.

The test was administered in a quiet room, free from distractions, but due to limited room availability at some school sites, in some instances, from two to five administrators conducted tests in a large spacious but undisturbed room, such as the cafeteria. This proved challenging at times, but students were still able to successfully complete testing in a satisfactory environment.

In most cases, all testing was completed between the times of 8:00 AM to 11:00 AM when the students were most attentive and the school sites tended to be more quiet and conducive to learning. During a designated time slot, students were individually called from class by an administrator and escorted to an assigned room away from the classroom. They were seated directly across from the administrator to create a formal testing atmosphere in the least distracting sitting arrangement possible.

Digital audio recorders were used during the administration of the test. Audio files were transcribed and revisited at a later time during scoring. It was decided that test administrators would not do any evaluations (rating, scoring, writing notes) at the time of testing to prevent the students from becoming nervous or apprehensive. A table of student names with numerical codes was provided and used to identify students in digital recordings while helping to maintain anonymity in file names.

### ***Rating Steps***

In order to assess oral language proficiency, the ‘Ōlelo Ola team had to develop a method for rating the collected data. The two form-focused, short oral response sections of the test were given a 1 for a correct response and a 0 for an incorrect response on the 45 items. Because the introduction and story-telling picture series tests used open-ended oral language prompts, they could not be rated in this manner, and therefore an assessment rubric was needed.

The ‘Ōlelo Ola team, with the guidance of experts on language assessment, collectively developed a Hawaiian oral language proficiency assessment rubric to rate the introduction and

story-telling picture series portions of the test. This rubric would represent Hawaiian oral language proficiency as defined by this team. See Appendix A (Hawaiian version of the rubric) and Appendix B (English version of the rubric).

The assessment rubric was developed with three levels of proficiency (novice, intermediate, and pre-advanced), in seven proficiency domains: communicative skill, vocabulary, grammar, pronunciation, fluency, language steadfastness, and cultural and linguistic authenticity.

Using a number of student examples from the data collection conducted in May 2009, after the initial interviews had been completed, the ‘Ōlelo Ola team and language assessment experts tested the initial draft of the rubric. Through experimentation, and resulting discussions, the team was able to improve the clarity and efficacy of the rubric, and build inter-rater reliability.

Students were rated on only the first two minutes of the introduction, and then rated on the entire story-telling picture series test. By listening to the audio file and using a transcription of these sections as reference, the rater used a 1-3 scale (where 1 = novice, 2 = intermediate, and 3 = pre-advanced) to rate students in each of the seven proficiency domains mentioned above. Each rater scored each of the students with whom they conducted the oral language test.

The entire introduction, story-telling picture series, and story-telling using the wordless book section, was later analyzed by all raters and used to create a personalized report for participating schools and teachers on the grammatical strengths and weaknesses of their school. Schools were also shown a box and whisker plot revealing the level of proficiency demonstrated by the students in their school in comparison to the students in other participating schools. The anonymity of the other schools was maintained in all documents.

## **RESULTS**

This section presents the results of the study. The analysis of the oral language assessment data is broken down into several categories, which include: (a) the descriptive statistics of the open-ended long response test and the short response test, (b) an item analysis of individual subtests in the assessment, (c) correlational analyses, and (d) a multifaceted Rasch analysis.



### *Descriptive Statistics*

Table 1 shows the descriptive statistics for the total scores for the Open-Ended Oral Response sections (including the Introduction and either the Lei or Slipper series) and the form-focused Short Oral Response sections of the H-OLA assessment. Notice that 270 students took the Introduction and Short Oral Response Tests. The Lei and Slipper picture series tests were each taken by approximately half of the students: the Lei Series was taken by 137 examinees and the Slipper Series by 133. Note also that the means, medians, and midpoints are very similar for the Introduction, Lei Series, and Slipper Series with all of them falling between 13 and 14. Since these indicators of central tendency are based on a test with a total of 21 points possible, the scores are reasonably well centered. The high, low, range, and standard deviation indicate that the scores for the Introduction, Lei Series, and Slipper Series are also fairly widely dispersed around the central tendency with room for at least two standard deviations above and below the mean in all cases. All of this indicates that the distributions of these three measures were reasonably normal in shape. The Total Series combines the Introduction and Lei Series scores with the Introduction and Slipper scores (as though two forms of the test were combined). Naturally, these statistics (based on 270 students and 42 points possible) also indicate that the scores are well centered and dispersed. The reliabilities for the Introduction, Lei Series, and Slipper Series were moderate at .80, .74, and .69, respectively, meaning that the seven series category ratings taken together were 80%, 74%, and 69% reliable and 20%, 26%, and 31% unreliable, respectively.

The 45 items of the Short Oral Response sections were also completed by 270 students. The pattern of high to low for the indicators of central tendency (median = 36.00, mean = 33.90, and midpoint = 26.00) and the fact that there is not room for two standard deviations above the mean in the distribution indicate that the distribution is somewhat skewed. In this case, it appears that this section of the test is a bit too easy for these students. Put another way, the short response items would probably function better if some of the easy items were eliminated and more difficult items were added. The reliability for the form-focused Short Oral Response section of the test turned out to be a moderately high .87 meaning that the 45 items taken together were 87% reliable and 13% unreliable.

Table 1  
*Descriptive Statistics for the Introduction, Lei Series, Slipper Series, Short Oral Response Test Totals*

Statistic	Introduction	Lei Series	Slipper Series	Total for Open-Ended Oral Response Items	Total for Short Oral Response Items
Number	270	137	133	270	270
Total Possible	21	21	21	42	45
Mean	13.25	13.46	13.83	26.89	33.90
Median	13.00	14.00	14.00	27.00	36.00
Midpoint	14.00	13.50	13.50	27.00	29.00
High	21	20	19	40	45
Low	7	7	8	14	13
Range	15	14	12	27	33
Standard Deviation	2.93	2.41	2.30	4.82	6.62
Reliability	.80	.74	.69	NA	.87

Tables 2, 3, and 4 show the descriptive statistics in a different way for the Introduction, Lei Series, and Slipper Series, respectively. Here the statistics are given separately for each of the seven categories that raters were scoring. The categories were Communicative Skill (Com), Vocabulary (Voc), Grammar (Gra), Pronunciation (Pro), Fluency (Flu), Steadfastness (Ste), and Cultural Authenticity (Cul). The means in the tables can be used to determine which categories on which series were scored highest and lowest. For example in Table 2, Steadfastness (Ste) was scored the highest with a mean of 2.41, and Cultural Authenticity (Cul) was scored lowest with a mean of 1.27. Interestingly, the pattern of highest and lowest categories is the same in Tables 3 and 4. In any case, such comparisons of categories may help in deciding, on the basis of difficulty, which categories to keep and which to abandon.

Table 2

*Descriptive Statistics for the Introduction Categories, Totals, and Averages*

Statistic	Com	Voc	Gra	Pro	Flu	Ste	Cul	Total	Average
Number	270	270	270	270	270	270	270	270	270
Mean	2.01	1.84	1.69	2.27	1.77	2.41	1.27	13.25	1.89
Median	2.00	2.00	2.00	2.00	2.00	3.00	1.00	13.00	1.86
Midpoint	2.00	2.00	2.00	2.00	2.00	2.00	2.00	14.00	2.00
High	3	3	3	3	3	3	3	21	3
Low	1	1	1	1	1	1	1	7	1
Range	3	3	3	3	3	3	3	15	3
Standard Deviation	.79	.63	.60	.53	.56	.67	.54	2.93	.42

Table 3

*Descriptive Statistics for the Lei Series Categories, Totals, and Averages*

Statistic	Com	Voc	Gra	Pro	Flu	Ste	Cul	Total	Average
Number	137	137	137	137	137	137	137	137	137
Mean	2.23	1.85	1.70	2.25	1.73	2.51	1.20	13.46	1.92
Median	2.00	2.00	2.00	2.00	2.00	3.00	1.00	14.00	2.00
Midpoint	2.00	2.00	2.00	2.00	2.00	2.00	2.00	13.50	1.93
High	3	3	3	3	3	3	3	20	3
Low	1	1	1	1	1	1	1	7	1
Range	3	3	3	3	3	3	3	14	3
Standard Deviation	.65	.55	.53	.51	.51	.64	.42	2.41	.34

Table 4

*Descriptive Statistics for the Slipper Series Categories, Totals, and Averages*

Statistic	Com	Voc	Gra	Pro	Flu	Ste	Cul	Total	Average
Number	133	133	133	133	133	133	133	133	133
Mean	2.29	2.05	1.74	2.24	1.83	2.52	1.15	13.83	1.98
Median	2.00	2.00	2.00	2.00	2.00	3.00	1.00	14.00	2.00
Midpoint	2.00	2.00	2.00	2.00	2.00	2.00	1.50	13.50	1.93
High	3	3	3	3	3	3	2	19	3
Low	1	1	1	1	1	1	1	8	1
Range	3	3	3	3	3	3	2	12	3
Standard Deviation	.66	.57	.53	.55	.56	.60	.36	2.30	.33

The high and low statistics in all three tables are all 3 and 1, respectively, indicating that the full range of possible points was used by the raters in doing the scoring. The standard deviations indicate which categories spread the students out most and least. For instance, in Table 2 the Communicative Skill (Com) category had the highest standard deviation of .79, while Pronunciation (Pro) had the lowest at .53. Such comparisons of the standard deviations for

categories may help in deciding, on the basis of score variation, which to keep and which to abandon.

Table 5 gives similar information about the six subtests on the Short Oral Response (SOR) sections of the test. Notice that all 270 students completed the 45-items of these sections of the test. The high statistic supports the fact that the first three subtests had six items in each and the last three subtests had nine items in each. Note also that some students answered all six or all nine items correctly in each subtest. The means for the Noun, Locative, Marker, Verb, and Pronoun subtests were all fairly high (being at least 2/3rds of the items in each case). The fact that the means, medians, and midpoints vary considerably for the Total SOR scores and the fact that there is not room for two full standard deviations above the mean, probably indicate that there are some problems with normality in the distribution of total scores as well as in most of the subtests. Clearly, Table 5 presents the statistics of a test (and subtests) in need of revision.

Table 5  
*Descriptive Statistics for the Short Oral Response Subtests and Totals*

Statistic	Noun	Dem. Pronoun	Locative	Marker	Verb	Pronoun	Total SOR
Number	270	270	270	270	270	270	270
Mean	5.29	3.54	4.47	6.57	7.34	6.70	33.90
Median	6.00	4.00	4.00	8.00	8.00	7.00	36.00
Midpoint	4.00	3.00	3.00	4.50	5.50	4.50	29.00
High	6	6	6	9	9	9	45
Low	2	0	0	0	2	0	13
Range	5	7	7	10	8	10	33
Standard Deviation	.89	1.56	1.14	2.79	1.40	1.80	6.62

### *Item Analysis*

The best strategy to use in revising all of these tests and making them function more efficiently is to conduct item analysis. Here we will examine the item statistics for the original versions of the Introduction, Lei Series, and Slipper Series tests as well as the original Short Oral Response Test, in terms of which categories/items should be kept and which could be deleted. Much later in the report, we will consider what revised, shorter versions of all these tests would probably look like statistically if they were administered to the same sorts of students.

Table 6  
*Item Statistics for the Original Versions of the Introduction, Lei, and Slipper Series*

Item/Subtest Description	Mean	<i>R</i>	<i>SD</i>	<i>SD</i> <sup>2</sup>
Introduction: Communicative Skill	2.01	.76	.79	.62
Introduction: Vocabulary	1.84	.77	.63	.39
Introduction: Grammar	1.69	.79	.60	.36
Introduction: Pronunciation	2.27	.53	.53	.28
Introduction: Fluency	1.77	.79	.56	.32
Introduction: Language Steadfastness	2.41	.49	.67	.45
Introduction: Cultural Authenticity	1.27	.59	.54	.29
Introduction: Total (Rubric Scores)	13.25	<b>.43</b>	2.93	8.59
Introduction: Average of Rubric Scores	1.89		.42	.18
Lei Series: Communicative Skill	2.23	.65	.65	.42
Lei Series: Vocabulary	1.85	.72	.55	.30
Lei Series: Grammar	1.70	.72	.53	.28
Lei Series: Pronunciation	2.25	.55	.51	.26
Lei Series: Fluency	1.73	.72	.51	.26
Lei Series: Steadfastness	2.51	.55	.64	.41
Lei Series: Cultural Authenticity	1.20	.50	.42	.17
Lei Series: Total (Rubric Scores)	13.46	<b>.58</b>	2.41	5.80
Lei Series: Average of Rubric Scores	1.92		.34	.12
Slipper Series: Communicative Skill	2.29	.64	.66	.43
Slipper Series: Vocabulary	2.05	.76	.57	.32
Slipper Series: Grammar	1.74	.71	.53	.28
Slipper Series: Pronunciation	2.24	.51	.55	.30
Slipper Series: Fluency	1.83	.65	.56	.32
Slipper Series: Steadfastness	2.52	.49	.60	.35
Slipper Series: Cultural Authenticity	1.15	.34	.36	.13
Slipper Series: Total (Rubric Scores)	13.83	<b>.51</b>	2.30	5.27
Slipper Series: Average of Rubric Scores	1.98		.33	.11
Intro & Series Prompts Total	26.89		4.82	23.23
Intro & Series Prompts Average	1.92		.34	.12

**Original versions of the Introduction, Lei Series, and Slipper Series.** Table 6 shows the means, correlation coefficients (*r*), standard deviations (*SD*), and variances (*SD*<sup>2</sup>) separately for each of the seven rating categories in the Introduction, Lei Series, and Slipper Series. The means tell us about the relative difficulty (or severity) of the ratings given in each category. The correlation coefficients (*r*) indicate the degree to which each category is related to the total scores for each measure, or put another way, these coefficients indicate how well each category spreads the students out relative to the way the total scores spread them out. The standard deviation (*SD*) and variance (*SD*<sup>2</sup>) are two slightly different ways of looking at how much scores varied in each category.

What does all of this mean? To begin with, notice that the lowest means in all three sets of seven rating categories are for the Cultural Authenticity category at 1.27, 1.20, and 1.15, respectively, and that the highest means are for the Language Steadfastness category at 2.41, 2.51, & 2.52, respectively. This simply means that the raters were consistently giving their

lowest ratings on average for Cultural Authenticity and their highest for Language Steadfastness. This pattern is interesting in itself, but taken together with the rest of the means, it seems that the raters were using these categories (low to high) in similar ways relative to each other in each of the three sets of seven rating categories.

The correlation coefficients indicate the degree to which each rating category spread the students out in a manner similar to the total scores for that set of seven categories. Because more observations are most often more reliable than fewer, the total scores on any test are taken to be more reliable than any single item or scoring category that contributes to that score. For this reason, we calculated the category/total correlation coefficients as an indication of how well each category is discriminating among the students in the same way that the total scores discriminated. We can use such information in trying to decide which categories we might want to eliminate in future versions of the test. For example, if we wanted to trim the number of categories used and thereby make the rating job easier, we might want to eliminate the lowest discriminating categories (Pronunciation, Steadfastness, and Cultural Authenticity) which discriminated as follows: for the Introduction, .53, .49, and .59, respectively; for the Lei Series, .55, .55, and .50, respectively; and for the Slipper Series, .51, .49, and .34.

Notice also that the same statistics are given for total scores and averages in each case. In addition, three of the correlation coefficients ( $r$ ) are in bold-faced type. (These are the correlations for the total scores on each of the Introduction, Lei Series, and Slipper Series with the Introduction combined with whichever Series each student took). These indicate that the total Introduction scores correlated somewhat at .43, the total Lei Series scores a bit better at .58, and the total Slipper Series scores less and better than the other two at .51.

**Original Short Oral Response.** Table 7 also shows the means, correlation coefficients, standard deviations, and variances separately for each item of the Short Oral Response sections of the test that all examinees took. Notice that the items are organized into six items each (and subtotals) for Nouns, Demonstrative Pronouns, and Locatives, as well as nine items each (and subtotals) for Verb Markers, Verbs, and Pronouns. The means for each item tell us about the relative difficulty (or severity) of the items. The correlation coefficients indicate the degree to which each item is related to the total scores for each subtest, or put another way, these coefficients indicate how well each item spreads the students out relative to the way the subtest

scores spread them out. The standard deviation and variance are two slightly different ways of looking at item variance.

What does all of this mean? To begin with, notice that the lowest mean is Verb: Wave = .19, which can be interpreted in this case (where the scoring is right/wrong) as the item facility, or proportion of examinees who answered correctly. By moving the decimal point two places to the right, we can interpret this as a percent (in this case, .19 becomes 19%). This means that, in this case, 19% of the examinees answered correctly. In other words, this was a difficult item for these examinees. After all, 81% got it wrong. However, two of the items had means (or item facility values) of .99 (Locative: Inside of and Verb: Sleep), which means that these two items were very easy for the examinees with 99%, or virtually everybody, answering them correctly. Good items for a placement or proficiency test are those that have item facility values around .50, say from .30 to .70.

Items in that range will also typically have relatively high point-biserial correlation coefficient ( $r_{pbi}$ ). The correlation coefficients in this case (where the data were coded 0 for wrong and 1 for right) are point-biserial coefficients. In situations like this these can also be called discrimination indexes. Whatever name they are given, these values indicate the degree to which each item is spreading the students out in a manner similar to the subtest scores. We can use such information in trying to decide which items we might want to eliminate in future revised versions of the test. For example, if we wanted to trim the number of items in each subtest to five (making the test 30 items long instead of 45 items long, we could eliminate the one item with the lowest  $r_{pbi}$  value in each of the Nouns, Demonstrative Pronouns, and Locatives subtests, as well as the four items with the lowest  $r_{pbi}$  value in each of the Verb Markers, Verbs, and Pronouns subtest. The resulting 30-item revised version of the test should not only be 15 items shorter than the original 45-item test, but also equally or more reliable than the .87 reliability found for this original version. Thus, the revised test should be much more efficient than the original version—being equally reliable, but quicker to administer.

Table 7  
 Item Statistics for the Original Version of the Short Oral Response Test

Item/Subtest Description	Mean	$r_{phi}$	SD	SD <sup>2</sup>
Noun: shell lei	.88	.52	.32	.10
Noun: spider	.90	.55	.30	.09
Noun: octopus	.91	.49	.28	.08
Noun: spoon	.90	.43	.29	.09
Noun: shark	.77	.49	.42	.18
Noun: elephant	.93	.33	.26	.07
<i>Noun: Subtotal</i>	<i>5.29</i>	<i>.48</i>	<i>.89</i>	<i>.80</i>
Demonstrative Pronoun: kēlā	.73	.29	.44	.20
Demonstrative Pronoun: kēia	.51	.62	.50	.25
Demonstrative Pronoun: kēnā	.54	.62	.50	.25
Demonstrative Pronoun: kēlā	.72	.38	.45	.20
Demonstrative Pronoun: kēnā	.47	.63	.50	.25
Demonstrative Pronoun: kēia	.55	.66	.50	.25
<i>Demonstrative Pronoun: Subtotal</i>	<i>3.54</i>	<i>.53</i>	<i>1.56</i>	<i>2.42</i>
Locative: On the side of	.62	.54	.49	.24
Locative: On top of	.95	.27	.22	.05
Locative: Behind	.41	.72	.49	.24
Locative: Under	.97	.28	.17	.03
Locative: Inside of	.99	.35	.12	.01
Locative: In front of	.53	.75	.50	.25
<i>Locative: Subtotal</i>	<i>4.47</i>	<i>.48</i>	<i>1.14</i>	<i>1.29</i>
Verb Marker: Ke...nei	.85	.73	.36	.13
Verb Marker: Ke...nei	.77	.80	.42	.17
Verb Marker: Ke...nei	.74	.79	.44	.19
Verb Marker: Ke...nei	.72	.75	.45	.20
Verb Marker: Ke...nei	.80	.81	.40	.16
Verb Marker: Ke...nei	.83	.81	.37	.14
Verb Marker: Ke...nei	.64	.75	.48	.23
Verb Marker: Ua	.64	.55	.48	.23
Verb Marker: E...ana	.56	.53	.50	.25
<i>Verb Marker: Subtotal</i>	<i>6.57</i>	<i>.84</i>	<i>2.79</i>	<i>7.78</i>
Verb: Laugh	.93	.42	.26	.07
Verb: Paddle (Canoe)	.77	.67	.42	.18
Verb: Sleep	.99	.25	.10	.01
Verb: Converse (on the telephone)	.93	.48	.26	.07
Verb: Cook	.86	.59	.35	.12
Verb: Eat / Picnic	.96	.33	.20	.04
Verb: Wave	.19	.50	.39	.15
Verb: Clap	.90	.53	.29	.09
Verb: Build	.83	.60	.38	.14
<i>Verb: Subtotal</i>	<i>7.34</i>	<i>.76</i>	<i>1.40</i>	<i>1.96</i>
Pronoun: they (2)	.81	.62	.39	.16
Pronoun: they (3+)	.93	.56	.26	.07
Pronoun: he/she	.86	.58	.34	.12
Pronoun: they (2)	.73	.62	.44	.20
Pronoun: he/she	.90	.47	.30	.09
Pronoun: they (3+)	.93	.41	.26	.07
Pronoun: you	.70	.60	.46	.21
Pronoun: us (2)	.22	.40	.42	.17
Pronoun: I	.62	.51	.49	.24
<i>Pronoun: Subtotal</i>	<i>6.70</i>	<i>.78</i>	<i>1.80</i>	<i>3.26</i>
Short Oral Response Test: Total	33.90		6.62	43.85



Notice also that the same statistics are given for the six sets of subtest total scores (in italics). In addition, six of the correlation coefficients are in bold-faced italics type. These are the correlations for the subtest scores with the total test scores. These indicate that the subtests correlate with the total scores as follows: Noun/Total and Locative/Total are low at .48, Demonstrative Pronoun/Total is a bit better at .53, while Verb Marker/Total, Verb/Total, and Pronoun/Total are all moderately correlated at .84, .76, and .78., if we decide to cut the number of subtests, any of the first three subtests might be candidates for elimination.

### *Correlational Analyses*

Table 8 shows all possible correlation coefficients for the Introduction, Lei Series, Slipper Series, Introduction and Series combined, as well as for the Noun, Demonstrative Pronouns, Locative, Verb Marker, Verb, Pronoun, and Total Short Oral Response Test scores. Notice that all but three are significant at  $p < .01$ . This simply means that there is only a 1% probability that these coefficients occurred by chance alone. However, the degree to which the individual correlation coefficients are interesting is a separate issue. Notice for instance that the Introduction, Lei Series, Slipper Series, as well as the Introduction & Series combination all correlate with each other at between .637 and .929. Thus these correlations indicate that the various sets of scores go together somewhat to a great deal, depending on the pairing involved. Put another way, the overlapping variances between the sets of scores ranged from 40.6% overlap to 86.3%, as indicated by  $r^2$  (i.e., where  $r = .637$  above,  $r^2 = .405769 \approx .406$ , a proportion equivalent to 40.6%; similarly, where  $r = .929$  above,  $r^2 = .863041 \approx .863$ , a proportion equivalent to 86.3%). For ease of interpretation, the squared values for each correlation coefficient are shown in Table 9.

Table 8

*Pearson Product-Moment Correlation Coefficients (Extended and Short Oral Response Sections)*

	Intro	Lei Series	Slipper Series	Intro & Series	Noun	Dem. Pronoun	Locative	Verb Marker	Verb	Pronoun	Total SOR
Intro	1.00	.637*	.673*	.929*	.202*	.132	.258*	.337*	.366*	.404*	.432*
Lei Series		1.00	<sup>a</sup>	.886*	.434*	.224*	.332*	.410*	.585*	.481*	.577*
Slipper Series			1.00	.891*	.093	.139	.289*	.459*	.390*	.456*	.513*
Intro & Series				1.00	.263*	.169*	.310*	.417*	.463*	.475*	.531*
Noun					1.00	.110	.170*	.303*	.368*	.319*	.482*
Dem. Pronoun						1.00	.168*	.259*	.242*	.341*	.531*
Locative							1.00	.266*	.350*	.226*	.482*
Verb Marker								1.00	.560*	.552*	.837*
Verb									1.00	.552*	.764*
Pronoun										1.00	.783*
Total SOR											1.00

\* $p < .01$ <sup>a</sup> no overlapping data, that is, half the students took each series, but no students took both

Table 9

*Coefficients of Determination (Extended and Short Oral Response Sections)*

	Intro	Lei Series	Slipper Series	Intro & Series	Noun	Dem. Pronoun	Locative	Verb Marker	Verb	Pronoun	Total SOR
Intro	1.000	.406	.453	.863	.041	.017	.067	.114	.134	.163	.187
Lei Series		1.000	<sup>a</sup>	.785	.188	.050	.110	.168	.342	.231	.333
Slipper Series			1.000	.794	.009	.019	.084	.211	.152	.208	.263
Intro & Series				1.000	.069	.029	.096	.174	.214	.226	.282
Noun					1.000	.012	.029	.092	.135	.102	.232
Dem. Pronoun						1.000	.028	.067	.059	.116	.282
Locative							1.000	.071	.123	.051	.232
Verb Marker								1.000	.314	.305	.701
Verb									1.000	.305	.584
Pronoun										1.000	.613
Total SOR											1.000

<sup>a</sup> no overlapping data, that is, half the students took each series, but no students took both

In addition, the correlations shown in Table 8 for Verb Marker, Verb, and Pronoun subtest scores with the Total Short Oral Response (SOR) scores are moderately high at .837, .764, and .783, respectively. These results make sense given that the correlations of subtests within the Short Oral Response sections are themselves part of the total scores (and sets of numbers correlate with themselves perfectly, thus raising any correlation involving the set of numbers and another set of which they are part). But why are these three more highly correlated with the total scores than the Noun, Demonstrative Pronoun, and Locative are? Perhaps this difference occurs because the Noun, Demonstrative Pronoun, and Locative only have six items in each subtest, while the Verb Marker, Verb, and Pronoun subtests have nine items each; the latter three subtests are therefore contributing a larger proportion of the variance to the total scores than the former. Alternatively, the Verb Marker, Verb, and Pronoun subtests appear to be somewhat more highly related to each other, correlating at between .552 and .560 with each other, while the correlations

involving the Noun, Demonstrative Pronoun, and Locative subtests are lower ranging from .170 to .368; this could lead to the Verb Marker, Verb, and Pronoun subtests working more closely together in their contribution to the total scores, which would in turn probably increase their correlation with those total scores.

### ***Multifaceted Rasch Analysis***

Multifaceted Rasch analysis (FACETS) is used to examine the degree to which variables and levels of those variables produce different scores relative to each other—all on the same scale called a logit scale. In this study, we were interested in the degree to which different raters are severe or lenient and rating categories are difficult or easy.

Table 11

*Preliminary Results for the FACETS Analysis for Introduction, Lei Series, Slipper Series, and Short Oral Response Items for Appropriate Facets*

<b>TEST</b> Facet	# Misfit	RMSE	Separation	Reliability	Chi-square (fixed)
<b>Introduction</b>					
Examinees	7	.89	2.50	.82	$p = .00$
Raters	0	.12	.25	.06	$p = .36$
Categories	0	.14	13.59	.99	$p = .00$
<b>Lei Series</b>					
Examinees	6	.91	1.85	.77	$p = .00$
Raters	0	.18	1.28	.70	$p = .02$
Categories	0	.21	11.79	.99	$p = .00$
<b>Slipper Series</b>					
Examinees	5	.87	1.70	.74	$p = .00$
Raters	0	.17	.00	.00	$p = .53$
Categories	0	.21	11.56	.99	$p = .00$
<b>Short Oral Response Test</b>					
Examinees	13	.49	2.21	.83	$p = .00$
Items	0	.22	6.67	.99	$p = .00$

**6.4.1 Preliminary results.** Table 11 shows the preliminary results for the four FACETS analyses conducted here (one each for the Introduction, Lei Series, Slipper Series, and Short Oral Response items). Notice that labels are given for five statistics across the top of the table: # Misfit, RMSE, Separation, Reliability, and Chi-square (fixed). Also notice in the first column that the rows are labeled with the four tests (along with the facets that are appropriate for each). Let's consider each of the statistics in turn.

The # *Misfit* indicates how many examinees, raters, categories, or items “did not fit the general pattern of responses in the matrix, and can thus be classified as relatively misfitting...”

(McNamara, 1996, p. 171). Notice that there were 7, 6, 5, and 13 misfitting examinees for the Introduction, Lei Series, Slipper Series, and Short Oral Response items, respectively, and that there were no misfitting raters, categories, or items. All of this means that 7, 6, 5, and 13 students were not fitting the measurement model in this analysis due to response patterns that were not expected in all but two of the cases. Two of the students completing the Short Oral Response test items received perfect scores of 45, which in Rasch analysis means that the test was not appropriate for these two students because it could not estimate if they were higher than the 45 total possible.

*RMSE* stands for *root mean square standard error* (for all non-extreme measures). The *RMSE* is used to calculate the separation index discussed in the next paragraph. However, it also serves as an estimate of standard error. The lower the *RMSE* the better the data fit the measurement model. The *RMSE* values in Table 11, ranging from .12 to .91, are relatively high indicating that none of these facets are fitting the model as well as might be desired.

The *separation* index tells us the degree to which each facet spreads the examinees, raters, categories, or items relative to their precision (Linacre, 2008, p. 149). The higher the value is, the more each facet is spreading the elements that it includes. Notice that the separation indexes for categories and items tend to be higher than the other values and that the rest are relatively low. All of this indicates that the categories and items facets are high in terms of the spread of the estimates relative to their precision, while the other facets are not.

The *reliability* estimates shown in Table 11 would more accurately be labeled *separation reliabilities*. According to Linacre (2008, p. 217):

This shows how reproducibly different the measures are. This may or may not indicate how “good” the test is in other respects. High (near 1.0) person and item reliabilities are preferred.

This “separation” reliability is somewhat the opposite of an interrater reliability, so low (near .0) judge and rater reliabilities are preferred.

For example, a high reliability for examinees indicates that the examinees consistently differ from each other, which is generally viewed as “good” from a norm-referenced testing viewpoint. In contrast, high reliability for raters would not typically be viewed as “good” because it indicates that the raters are consistently different from each other in the severity or leniency of the ratings they assign. The degree to which consistent differences among categories or items is important is a different sort of issue. From our point of view, there is no problem if one category

is consistently scored lower or higher than the others, or if some items are consistently lower or higher than others. All in all, these reliability estimates should be interpreted as just what they are: estimates of the degree to which the test is consistently separating its examinees, raters, categories, or items. In Table 11, all measures appear to be reasonably reliable or consistent with regard to examinees. These estimates are similar to the more familiar Cronbach alpha estimates of reliability that are reported elsewhere in this report. The estimates in Table 11 for raters are zero in two cases and .70 in another. This means that raters do not vary consistently from each other in two cases, but do vary with moderately consistency on the Lei Series. Category and Item reliabilities are .99 in all cases, meaning that the difficulties of categories and items are consistently varying from each other.

The *chi-square (fixed)* statistic tests the following hypothesis: “Can this set of elements be regarded as sharing the same measure after allowing for measurement error?” Thus for the facets in this design, the following hypotheses are being tested:

1. Can these examinees be thought of as equally able?
2. Can these raters be thought of as equally severe or lenient in their ratings?
3. Can these categories be thought of as equally difficult?
4. Can these items be thought of as equally difficult?

The chi-square statistics in this study were found to be significant (at  $p < .01$ ), except for the raters’ facet on the Introduction, Lei Series, and Slipper Series. We cannot reject the null hypothesis that these raters are equally severe or lenient in their ratings. Hence we can only accept the hypothesis the raters are giving ratings of similar severity or leniency. All of the other hypotheses should be rejected, that is, the differences between examinee, categories, and items can be said to be statistically significant.

**6.4.2 Vertical rulers.** Next we will display and interpret the vertical rulers from our FACETS analyses. Four of these are shown in total (see Figures 1a to 1d) for the Introduction, Lei Series, Slipper Series, and Short Oral Response items, respectively.

Focusing first on the oral tests, notice that the first column in the vertical ruler for the Introduction, Lei Series, and Slipper Series (Figures 1a-1c) is for measure, which represents the range of scores on a true interval logit scale where the mean is 0 and, in this case, the range is -/+ 5 or 6. The second column shows where the examinees were on the scale (with each asterisk equivalent to 2 or 3 examinees as labeled at the bottom of that column in each figure). The third

column gives the averages for each of the five raters (R1, R2, R3, R4, & R5). The fourth column shows that the average ratings for each of the seven categories: Communicative Skill (Com), Vocabulary (Voc), Grammar (Gra), Pronunciation (Pro), Fluency (Flu), Steadfastness (Ste), and Cultural Authenticity (Cul). The final column shows the raw score equivalents along the same scale.

Figure 1a. Vertical Ruler for the Introduction

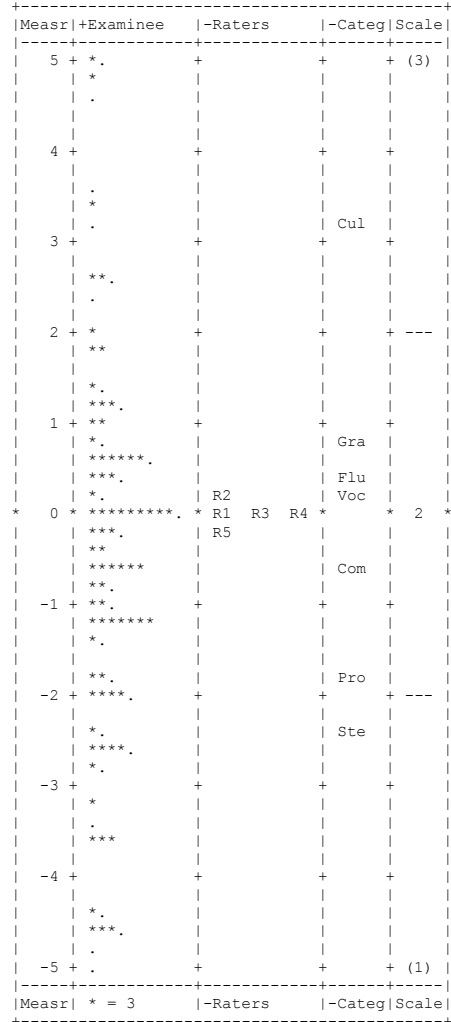


Figure 1a is the vertical ruler for the Introduction ratings. Notice that the examinees' scores range from -5 logits for the low scorers to +5 logits for the high scorers. This indicates reasonably wide differences in the performances/ratings of the students involved in this project. The third column shows that R2 was the most severe rater, and R5 was the most lenient rater with R1, R3, and R4 in between, but these differences were very small. The fourth column shows

that the average ratings for Cultural Awareness (Cul) were suitable for students who scored slightly above +3 logits (i.e., the ratings were low on average). In stark contrast, the average ratings for Language Steadfastness (Ste) ratings were suitable for students who scored midway between -2 and -3 logits (i.e., the ratings were high on average). The other five categories in descending order of suitability for high scoring examinees to low were: Grammar (Gra), Fluency (Flu), Vocabulary (Voc), Communicative Skill (Com), and Pronunciation (Pro). The last column to the far right shows how the raw ratings 1, 2, and 3 matched up to the true interval logit scores on the far left for the Introduction ratings overall.

Figure 1b. Vertical Ruler for the Lei Series

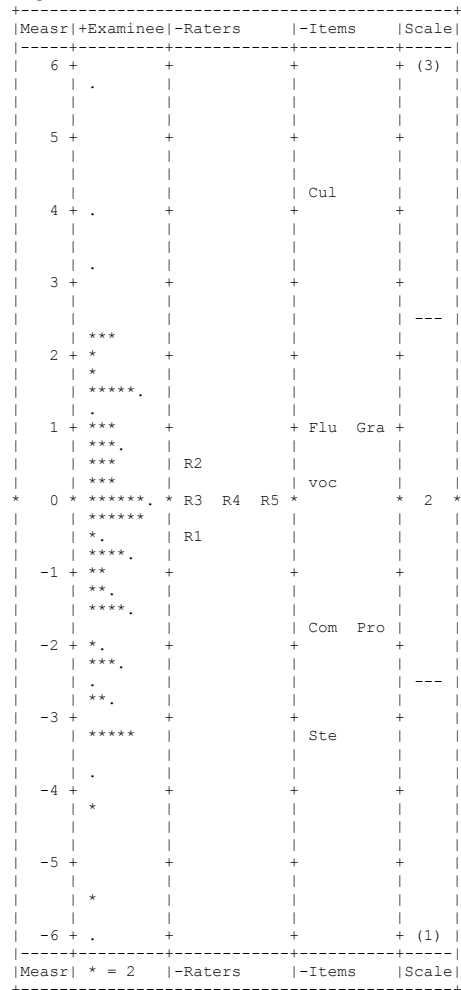


Figure 1b is the vertical ruler for the Lei Series ratings. Notice that the examinees' scores range a bit more than those in the previous figure from -6 logits for the low scorers to almost +6 logits for the high scorers. This indicates reasonably wide differences in the performances/ratings of those who took this series. The third column shows once again that R2 was the most severe

rater, but this time R1 was the most lenient with R3, R4, and R5 in between; these differences were a bit more substantial than those shown in the previous figure, but were still not very substantial. As with the previous figure, the fourth column shows that the average ratings for Cultural Awareness (Cul) were suitable for high scoring students, in this case those who scored somewhat above +4 logits (i.e., the ratings were low on average). In stark contrast and also similar to the previous figure, the average ratings for Language Steadfastness (Ste) were suitable for students who scored low, this time below -3 logits (i.e., the ratings were high on average). The other five categories in descending order of suitability for high scoring examinees to low were: Grammar (Gra), Fluency (Flu), Vocabulary (Voc), Communicative Skill (Com), and Pronunciation (Pro).<sup>2</sup> The last column to the far right shows how the raw ratings 1, 2, and 3 matched up to the true interval logit scores on to the far left for the overall Lei Series ratings.

Figure 1c is the vertical ruler for the Slipper Series ratings. Notice that the examinees' scores range less than those in the previous two figures from -5 logits to about +4 logits. This still indicates reasonably wide differences in the performances/ratings. The third column shows that R4 was the most severe rater, but this time R1 was the most lenient rater with R2, R3, and R5 in between;<sup>3</sup> these differences were small like those in Figure 1a. As in the previous two figures, the average ratings for Cultural Awareness (Cul) were suitable for high scoring students, in this case between +4 and +5 logits (i.e., the average ratings were low). Also similar to the previous two figures, the average ratings for Language Steadfastness (Ste) were suitable for students who scored low, this time slightly above -3 logits (i.e., the average ratings were high). The other five categories are in the same general descending order of high to low examinee suitability with the exception of Pro and Com (which have switched positions): Grammar (Gra), Fluency (Flu), Vocabulary (Voc), Pronunciation (Pro), and Communicative Skill (Com). The last column to the far right shows how the raw ratings 1, 2, and 3 matched up to the true interval logit scores on to the far left for the overall Slipper Series ratings.

---

<sup>2</sup> While Grammar and Fluency appear to be exactly the same in the Figure as do Communicative Skill and Pronunciation, the order described in text is reflected in small differences in the actual logit scores for these categories.

<sup>3</sup> While R2 and R4 appear to be exactly the same in the Figure as do R1 and R3, the order described in text is reflected in small differences in the actual logit scores for these pairs of raters.



Figure 1c. Vertical Ruler for the Slipper Series

Measr	Examinee	Raters	Items	Scale
5	+	+	+	(3)
			Cul	
4	+ *	+	+	
3	+ *	+	+	
2	+ *, *, *, * *****	+	+	---
1	+ *	+	+	
			Gra	
			Flu	
0	* *****	R2 R4 R5	* *	* 2 *
	**	R1 R3	Voc	
-1	+ ***** *, *, *, *, *, *, *	+	+	
			Pro Com	
-2	+ *	+	+	---
-3	+ *	+	+	
			Ste	
-4	+ *	+	+	
-5	+ *	+	+	(1)

Figure 1d. Vertical Ruler for the Short Oral Response Test

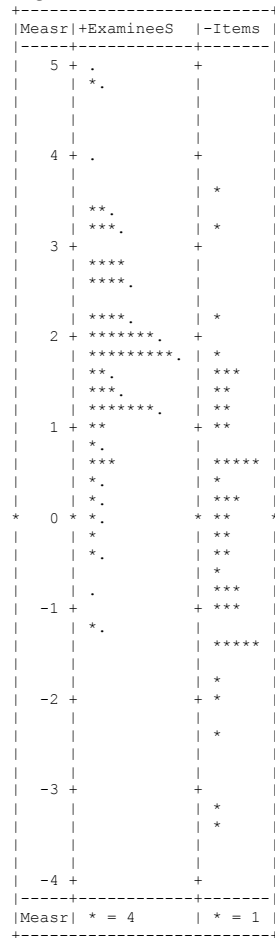
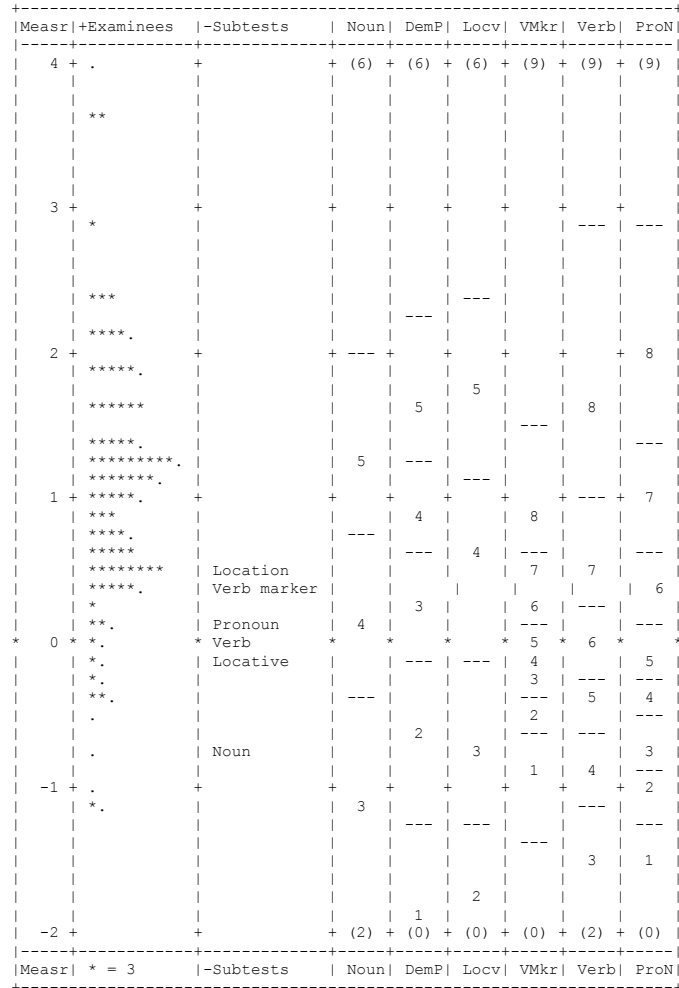


Figure 1d is the vertical ruler for the Short Oral Response items. Notice that the examinees' scores range less than those in the previous two figures from a bit lower than -1 logits to almost +5 logits. Because this test had no raters or categories, the only other column shown in the vertical ruler is for items. Notice that they range from a bit below -3 logits to a bit below +4. The mismatch between the examinee logits and item logits indicates that a number of the items were too easy for the examinees in this sample. More importantly there were no items difficult enough to be suitable for the high performing students.

Figure 1e is the vertical ruler for the Short Oral Response items that resulted from what is called a partial credit analysis. Notice that the subtests range in suitability for high scoring examinees to low in the following order (from midway between -1 and 0 to midway between 0 and +1) as follows: Demonstrative Pronoun, Verb Marker, Pronoun, Verb, Locative, and Noun. Notice also the six columns to the right, one each for each subtest. These show how the raw

scores on the subtests matched up with the logit scores. Notice how very different the suitability of the items was on each subtest for low to high scoring examinees.

Figure 1e. Vertical Ruler for the Short Oral Response Test (Partial Credit Model for subtests as units of analysis)



**6.4.3 Probability curves.** FACETS analysis also provides information in the form of probability curves. These tell us about the degree to which the points on the rating scale are separate or overlapping. The probability curves for the three long response oral tests in this study are shown in Figures 1a to 1c. To understand such graphs, we need to keep in mind that ideal probability curves have a distinct *hill-like* look with little overlap between curves, in this case, one each for the three scores, 1, 2, and 3. Notice in Figures 1a to 1c that the curves for the Introduction, Lei Series, and Slipper Series scores are all reasonably steep and hill-like with some overlap. What overlap there is appears to be due to heavy use of the 2 score by raters.

Figure 2a. Introduction Probability Curves

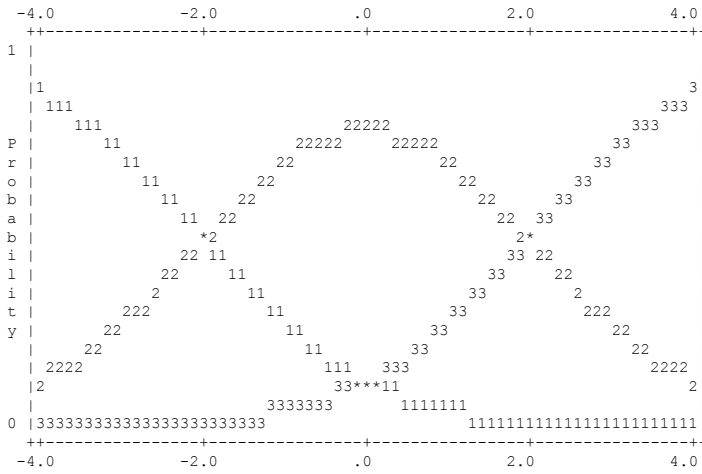


Figure 2b. Lei Series Probability Curves

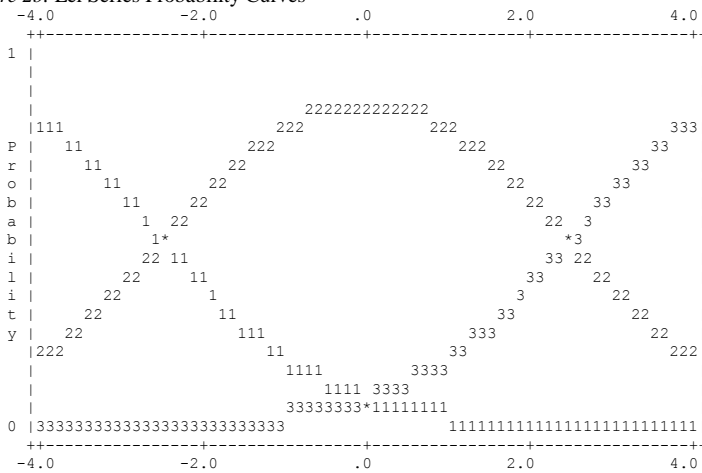
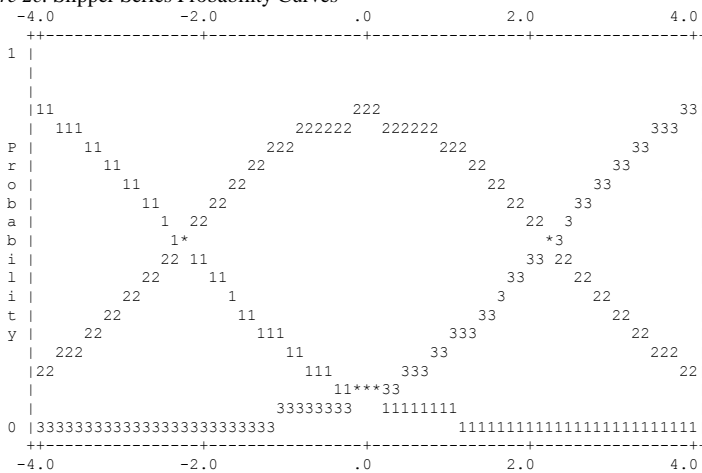


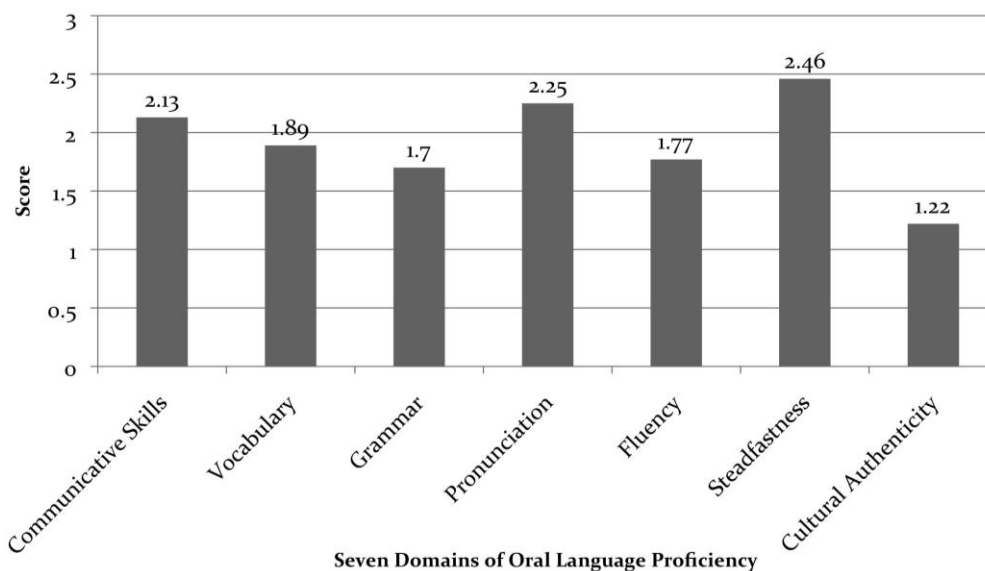
Figure 2c. Slipper Series Probability Curves



## DISCUSSION

This section includes a discussion of the results of the Hawaiian Oral Language Assessment. An analysis of the open-ended long response sections (Introduction, Lei Series, and Slipper Series) and the findings from the short oral response items will be presented. Figure 3 below shows the cumulative scores of the students' performance in the long response sections of the test in each of the seven proficiency domains: Communicative Skills, Vocabulary, Grammar, Pronunciation, Fluency, Steadfastness, and Cultural Authenticity. The seven domains will be addressed in order from the highest cumulative score to the lowest cumulative score, with the exception of the Grammar domain, which will be last. The Grammar domain is addressed last since the findings from the Long Response Test and the Short Oral Response items will be presented together in that section.

Figure 3. Cumulative Scores of Students in Grades 1-3 for Introduction and Picture Series



The results of the long response tests

(Introduction, Lei Picture Series, and the Slipper Picture Series) showed that students scored the highest in the Steadfastness domain. In order to score a three, students had to consistently use Hawaiian during the assessment period, which typically lasted about 20 minutes. A score of two means that students used Hawaiian the majority of the time, but interjected English words when they were not sure of the Hawaiian vocabulary. A student was given a one if she or he frequently used English during the prompt. The Steadfastness score is 2.41 for the Introduction, 2.51 for the

Lei Series, and 2.52 for the Slipper series. When these three scores are averaged, the resulting composite value is 2.46. It is interesting to note that the score for both the Lei Series and the Slipper Series are very close with a slight decrease for the Introduction. Perhaps the difference between the Introduction, Lei Series and Slipper Series prompts occurred since students followed the storyline that was evident in the sequence of the six pictures, whereas, during the Introduction section, students followed the basic sequence, but then also added more information regarding their home life and interests, perhaps using more vocabulary that they were not familiar with in Hawaiian. The evaluators of the oral language assessment found the Steadfastness domain to be a great strength of students in the Hawaiian language immersion program in grades 1-3. The result is an indication of a high level of commitment of students to maintain the Hawaiian language at all times and a demonstration of the high level of fluency that has been achieved in the early grades.

The second highest cumulative score was found in the Pronunciation domain. The score is 2.27 for the Introduction, 2.25 for the Lei Series, and 2.24 for the Slipper Series. All of these scores are very similar, demonstrating the consistency of pronunciation between the prompts. The average of the three scores is 2.25. To score a three, students must consistently pronounce *hakalama* (consonant-vowel clusters), vowel blends, *‘okina* (glottal stop), *kahakō* (macron), and phonemes correctly. A score of two means the student mispronounces some aspects of the items listed above. If pronunciation errors are frequent and obvious, a student will receive a score of one. Generally speaking, students did well in this domain. The most common mistake found is the insertion or deletion of the *‘okina* and the *kahakō*. Another area of weakness to be addressed is vowel blends, which is most likely due to interference from the students’ first language of English. Therefore, the basic rules for pronunciation mistakes should be explicitly taught and incorporated into lessons in the lower elementary grades, so that these types of errors don’t become fossilized and difficult to change as students grow older.

The third highest cumulative score was the Communicative Skills domain. In this category, students were given a three if complete sentences were used, ideas were expressed in a clear and easily understood manner, most aspects of the task were included in the student’s response, and communications were independently directed without relying on prompts or assistance from the evaluator. For a score of two, students sometimes spoke at the phrase level and sometimes at the sentence level, ideas expressed were mostly clear in meaning, some important aspects of the task

were included in the student's independent response, and assistance in producing responses was required by the student. To score a one, the student spoke in brief and incomplete sentences sometimes involving only one word or phrase, ideas were unclear, few aspects of the task were included in the student's independent response, and assistance and prompts were often required. The mean score for the Communicative Skills domain is 2.01 for the Introduction, 2.23 for the Lei Series, and 2.29 for the Slipper Series. The composite average score for this domain is 2.13. Again we see a close similarity between the two story-telling picture series, but a slight decrease in the Introduction score, suggesting perhaps that the task for presenting an independent introduction of oneself for two minutes is a slightly more difficult for younger children than looking at a sequence of pictures and then independently telling an story.

In the Vocabulary domain, the Introduction score is 1.84, the Lei Series score is 1.85 and the Slipper Series is 2.05 with a cumulative average score of 1.89. In this domain, the Introduction score and Lei Series score are more closely related than the Slipper Series score. The Introduction task deals with aspects of home life and interests at home, therefore students need to know vocabulary words that are typically used outside of the immersion school environment to successfully complete this task. The Lei Series also includes elements in the picture that take place in the community such as gathering flowers from a tree in the yard, putting flowers in a basket, making leis, driving to the airport, and giving leis to visiting grandparents. The setting of the Slipper Series, however, takes place at school with a situation that most young children experience: doing schoolwork in class, going to the playground when the recess bell rings, taking slippers off and leaving them on the side, playing kickball, returning to class when recess is over, being the last one to get your slippers and finding a pair of mix-matched slippers. The familiarity of the Slipper Series context and the likelihood of in-school Hawaiian language modeling and support for such a context is one possible explanation for a higher score in this domain. Since the majority of immersion students speak English at home, this creates an obstacle for teachers. Students tend to know *school vocabulary* very well because they use it everyday. However, when children need to use words that are more commonly used at home, evaluators discovered that many of the students didn't know those words, i.e. *wahī* (to wrap a present), *'ie* (basket), *kahua ho'olulu mokulele* (airport), and *mānai* (needle for string leis). On the positive side, when students didn't know particular vocabulary words, they did use strategies to communicate their ideas. Although some students said the word in English because they didn't know the equivalent

in Hawaiian, many others tried to find a similar word such as box or bag for the word *basket*, while others used circumlocution techniques to express their thoughts. Therefore, this challenge behooves teachers and parents to expose children to a variety of circumstances in which vocabulary can be learned. If we want students to be able to speak in Hawaiian in every context outside of school, vocabulary is one area that needs to be actively and explicitly taught to students.

In the Fluency domain, the Introduction score is 1.77, the Lei Series score is 1.73 and the Slipper Series is 1.83, with a composite average score of 1.77. All of the scores are somewhat similar, demonstrating consistency between the three separate tasks. This domain illustrates the students' proficiency in the phonological aspects of the language. These aspects include an ease and natural flow of speech in an authentically Hawaiian way, i.e., proper inflection, intonation, emphasis, rhythm, and appropriate pauses. The Fluency domain is another area of challenge for Hawaiian language immersion students for several reasons. First of all, outside of the Ni'ihau community, the students are rarely exposed to a native speaker. Secondly, almost all of the HLIP teachers are second language learners themselves and therefore may not be using authentic aspects of fluency in their own Hawaiian speech patterns. In addition, the students' first language of English often interferes with the correct flow of Hawaiian fluency. A glaring example that was heard over and over again during the interviews is the use of the word *um*. Native speakers will interject utterances such as 'Ō and 'Ā when thinking, but HLIP students have a tendency to fall back on English sounds. Another example is the rise of the voice in English at the end of a sentence when a question is asked. Many students tend to improperly use the same intonation when asking questions and even when making statements in Hawaiian. When HLIP schools first began, *kūpuna* (elders) were an integral part of the program. They provided excellent examples of native speech for the children to hear. Unfortunately, 23 years have passed since immersion began, so *mānaleo* (first-language Hawaiian speaking) *kūpuna* are often no longer available to work in HLIP schools. Since students don't have access to native speakers today, one idea is for teachers to use recordings of native speakers in their instruction so that students can hear and imitate authentic examples of proper fluency patterns.

Cultural Authenticity is the lowest scoring domain to be addressed. The Introduction score is 1.27, the Lei Series score is 1.20 and the Slipper Series is 1.15, with a composite average score of 1.22. In this domain, the students were evaluated on how well they used traditional features of



speech and communication in their language. The Cultural Authenticity domain was included in this study because evaluators wanted to see if students were using Hawaiian thought and perspectives in the construction of language. Students were scored on their ability to use oratorical features such as reciting *mo 'okū 'auhau* (genealogy), complementary pairs, opposites, idioms, famous sayings, proverbs, along with using culturally correct phrases such as, *Aia lākou ma luna o ke ka'a*. (They are **on** the car), rather than the incorrect phrase, *Aia lākou ma loko o ke ka'a*. (They are **in** the car). Generally speaking, it seems that students are committed to speaking in Hawaiian, but English thinking sometimes interferes with sentence construction. However, several students did score high in this domain, but appeared to be students from only a few classes. Therefore it appears that certain teachers make a special effort to focus on rich language experiences in the classroom, which include memorized verses, phrases, and wise sayings taken from traditional stories and chants. It was also evident that certain classes continue to focus on family lineage and more elaborate memorized introductions and build upon the introductions that are first taught at the Pūnana Leo Preschools.

The last domain is the Grammar domain. The Introduction score is 1.69, the Lei Series score is 1.70 and the Slipper Series is 1.74, with a cumulative average score of 1.71. Out of all seven domains, the results of the Grammar domain are the most consistent between the three tasks. The Grammar domain had the second to the lowest score in the oral language assessment and therefore is an area that needs attention. In addition to the statistical analysis that was conducted to get an overall general idea in terms of performance in each of the seven domains, an error analysis of grammar was conducted. Students' recordings were transcribed, analyzed, and compiled in an individualized report for each school that participated in the study. Codes were created to summarize the strengths and weaknesses of 12 overarching categories and 48 subcategories of grammar. Along with each summary report, printed examples of students' language coinciding with the 48 subcategories were also given to teachers at the reporting meeting at each school, with the intention that teachers would create lessons to address the specific weaknesses of their students in the area of grammar. The analysis and examples of grammar are very comprehensive and too long to include in this paper and should be addressed in a separate paper focused solely on grammar. However, summarized below is a small sample of findings from the Short Oral Response items of H-OLA. The graphs show cumulative scores of the students' performance in the short response sections. The test measured proficiency in six

different grammatical structures including nouns, demonstrative pronouns, locatives, verb markers, verbs, and pronouns.

Figure 4a. Cumulative Scores for Short Response (Six-Item Subsets)

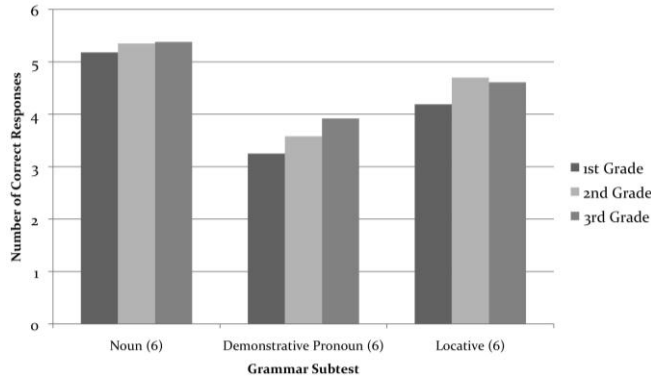


Figure 4b. Cumulative Scores for Short Response (Nine-Item Subsets)

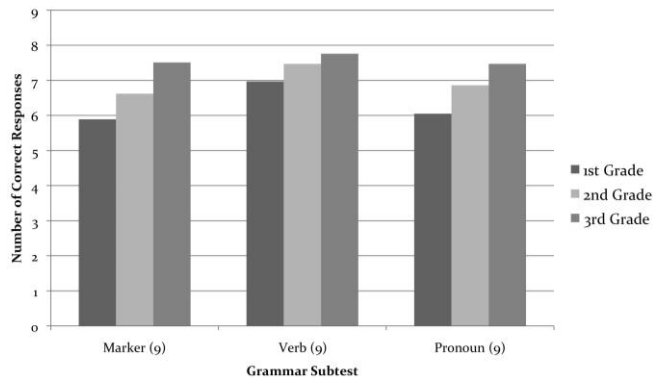


Figure 4a shows structures in which subsets of six items were tested, while Figure 4b shows structures in which subsets of nine items were tested. It is interesting to note that although five out of six of the categories show a natural progression from year to year, it appears that in the *locative* category, students in grade two performed slightly better than those in grade 3.

Students did very well on the Noun section, however, some students struggled with the Demonstrative Pronoun section. Students consistently used the word *kēia* correctly, but did not consistently use *kēnā* and *kēlā* correctly. The distinction between these two words for *that* is difficult for many second language speakers of Hawaiian, as it involves a different way of thinking about space and location as opposed to the English equivalent. In English, there is only one word for *that*, however, in Hawaiian, the speaker must distinguish between *that* (by you, the person being addressed) and *that* (far from you).

Students generally did well in the Locative section, with the exception of the words for *mua* and *hope*. The student sat on one side of a box and the test administrator sat on the opposite side

of the box. So when the test administrator placed an object either in front or in back of the box and asked where the object was in relationship to the box, the students sometimes gave an answer from the perspective of the student and sometimes gave an answer from the perspective of the test administrator. To alleviate the confusion, perhaps it would be best for the test administrator to sit next to the students instead of across from the student. It is possible that the scores for the Locative items were adversely impacted by this flaw in methodology as opposed to student ignorance of the meanings of these words.

Students did well in the Verb section, but results varied in the Verb Marker section from school to school. Students from certain schools knew the verb markers well, while students from other schools either dropped the markers altogether, dropped one half of the verb marker, or created interesting combinations of the different tense markers. Students from one particular school even created a new marker, *He verb nei*. The use of verb markers in the long response section was also analyzed to see if the findings were consistent with the short response section. However, evaluators found that students tended not to use verb markers in the introduction or the storytelling task. Therefore, verb markers are a weakness that needs to be looked at and addressed in the Hawaiian language immersion classroom.

In the Pronoun section, students know the singular pronouns (*au*, *'oe*, *'o ia*) well. However, confusion lies in the dual pronouns (*kāua*, *māua*, *'olua*, *lāua*). Many students seem to categorize pronouns into two areas, singular and plural, and therefore tend to overuse plural pronouns and underuse dual pronouns. The pronouns *kākou* and *lākou* especially are used as general terms for *we* and *they*. This is another indication of the English language interfering with Hawaiian thinking and perspectives, since there is only one word to express these thoughts in English, but there are several specific choices in Hawaiian.

In summary, the findings of the short oral response items of H-OLA manifest a natural progression in growth from Grade 1 to Grade 3. When all of the subtest scores of the short oral response items from all seven schools are combined together by grade level, students in Grade 1 demonstrate an average accuracy rate of 31.53 out of 45 items, which is 70% accurate. Students in Grade 2 show an average accuracy rate of 34.57 out of 45 items, which is 76.8% accurate. Students in Grade 3 display an average accuracy rate of 36.65 out of 45 items, which is 81.4% accurate.

## CONCLUSIONS

Overall, students in the Hawaiian language immersion program are performing very well in the Grades 1-3 levels. After assessing the proficiency levels of 270 students from seven different HLIP schools, the evaluators of the Hawai‘i Oral Language Assessment are confident to say that the obvious strengths of the program are: (a) the students’ steadfastness in using Hawaiian, (b) the correct pronunciation of words, and (c) and the high levels of communicative skills that are being demonstrated by the students in the early elementary grades. Students have the ability to express their thoughts through clear and descriptive Hawaiian without the need to code-switch to English or other languages. Therefore it appears that students are accomplishing one of the major goals of the program, which is to develop a high level of proficiency in comprehending and communicating in the Hawaiian language.

Students are also making progress in the areas of: (a) vocabulary, (b) grammar, and (c) phonological aspects of the language such as proper inflection, intonation, emphasis, rhythm, and appropriate pauses. However, the development of materials and the use of best-practice language acquisition teaching strategies, reinforced with excellent native speech examples and explicit instruction, would help to raise the proficiency levels of students in these three areas.

The greatest weakness that was found in the study is the area of cultural authenticity. Even though students are speaking in the Hawaiian language, this does not mean that they are automatically thinking and constructing language with Hawaiian thought and perspectives. This finding also has implications in other areas such as grammar for example. Many of the grammatical mistakes are due to the interference of the students’ first language of English, such as in the case of demonstrative pronouns and dual pronouns.

Students naturally follow the examples that are presented to them by their teachers and parent. Therefore, it is imperative that adult role models provide the foundation for Hawaiian thought and perspectives. The *Kumu Honua Mauli Ola* philosophy is one such model. There are other Hawaiian philosophical models that have also been developed by various groups. Whichever model is chosen, it is important for students to understand that Hawaiian thought plays an important role in perpetuating the traditional aspects of the language that have been passed down by ancestors for generations.

### ***Limitations of the Study***

The developers and evaluators of the Hawaiian Oral Language Assessment would have preferred to be able to assess every student at all 15 HLIP elementary schools throughout the state. However, due to restrictions on financial resources, human resources, and time, it was not possible. Given the above limitations, the assessment team carefully planned and followed statistical guidelines to assess a representative sampling of students from Hawaiian language immersion laboratory schools, charter schools, and regular immersion schools located on four major islands. Therefore, the team feels that enough data was collected to make generalizations regarding the progress of HLIP students in the early elementary grades.

The results of this study were collected from a 20-30 minute session between one student and one test administrator. It is important to keep in mind, that the interview session is only a small snapshot in a larger timeframe in the life of a HLIP student. If a student had a bad day, then perhaps that might reflect in the sample interview collected. However, since 270 students were assessed, the research team still feels they have a good sample from which to draw conclusions.

### ***Implications for Future Research***

Funding for this study was only sufficient to assess students in Grades 1-3. In 1989-1990 and 1990-1991, Warner conducted the first systematic study of HLIP students in Grades K-4. It would be interesting to take a closer in-depth look at grammatical structures that were produced in this assessment and compare it to Warner's study to determine the progress that has been made over the past twenty years.

Hawaiian language immersion program students beyond the fourth grade level have yet to be assessed and studied. In order to truly understand the language acquisition of students from Kindergarten to Grade 12, a comprehensive study should be developed and administered. Through systematic and longitudinal studies, language benchmarks can be determined and used to provide targets in continuous language growth.

Another area of research that can be addressed is assessing the proficiency levels of Hawaiian language immersion teachers. The purpose of such a study should be to help teachers and should not be used for punitive measures. In addition, if a study were to be conducted, follow-up classes should be made available to teachers to foster the growth of language skills in areas of need identified by the assessment.

### ***Final Thoughts***

This chapter only includes the results of the first assessment that was administered in April and May of 2009. The research team returned during the same months in 2010 to assess the same students (now representing Grades 2-4) for the second time. Since the scoring and statistical analysis of the second administration of the test is still ongoing, the longitudinal results of 2010 will be determined at a later time. Once the results of 2010 are complete, a comparison can be made between the two years of assessment.

As mentioned in the introduction of this paper, if the vision, mission, and goals of the Hawaiian language immersion program are to be realized, then several things need to transpire. For one, excellent language models need to be provided for children both at school and in the home. This can be accomplished by providing ongoing teacher training and parent classes or workshops. Secondly, additional materials both in print and non-print resources need to be developed and disseminated to schools and families. In order to produce language materials, funding needs to be made available. Lastly, assessment and evaluation for the purpose of improving instruction and raising the language proficiency levels of HLIP teachers and students needs to be ongoing. *E ola mau ka 'ōlelo Hawai'i* (May the Hawaiian language live on forever).

## REFERENCES

- ‘Aha Pūnana Leo, Inc. (no date). *A timeline of hawaiian language revitalization*. [Brochure]. Honolulu, HI: Author.
- Cherokee Nation. (2009). *Cherokee nation immersion school expands*. Retrieved on October 16, 2010 from <http://www.cherokee.org/NewsRoom/FullStory/3074/Page/Default.asp>
- Cherokee Nation. (2003). *Ga-du-gi: A vision for working together to preserve the Cherokee language*. Report on DHS ANA Grant #90-NL-0189. Tahlequah, Oklahoma: Cherokee Nation.
- Edmonds, C. (2008). The Reliability and Validity of the Māori language proficiency in writing test: Kaiaka Reo Year Eight. *Dissertation Abstracts International*, (UMI No. 3336175)
- Everly, H. (1965). *Education in Hawai‘i—Yesterday and today*. In *the Kamehameha Schools 75th anniversary lectures* (pp. 45-52). Honolulu: Bishop Museum Press.
- Hale Kuamo‘o-University of Hawai‘i at Hilo. (2010). [*Nā kula kaiapuni Hawai‘i 2010-2011*] Unpublished raw data.
- Hawai‘i General Superintendent of the Census. (1897). *Report of the general superintendent of the census, 1896*. Honolulu, HI: Hawaiian Star Press.
- Ka Haka ‘Ula O Ke‘elikōlani-University of Hawai‘i at Hilo. (2007). ‘Ōlelo ola oral language proficiency project. A grant proposal under the Native Hawaiian Education Program, US Department of Education, CFDA Number: 84.362A.
- Kana‘iaupuni, S. M. & Ishibashi, K. (2003, June). *Left behind: The status of Hawaiian students in Hawai‘i public schools*. Honolulu, HI: Kamehameha Schools, PASE Report 02-03:13.
- Kawai‘ae‘a, K., Housman, A., & Alencastre, M. (2007). Pū‘ā i ka ‘ōlelo, ola ka ‘ohana: Three generations of Hawaiian language revitalization. *Hūlili: Multidisciplinary Research on Hawaiian Well-being*, Vol. IV, 183-237.
- Linacre, J. M. (2008). *User’s guide to FACETS: Rasch-model computer programs*. Chicago: Author.
- Littler, R. (2001). *Report on Kaiaka Reo Data*. New Zealand: University of Waikato.
- Linn, M. (2004) Oklahoma’s native languages. Retrieved September 15, 2010 from <http://www.snomnh.ou.edu/collections-research/cr-sub/nal/ok%20spkrs%20w-notes.pdf>.
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.

- Office of Instructional Services, General Education Branch. (1994). *Long-range plan for the Hawaiian language immersion program: Papahana kaiapuni Hawai'i*. Hawai'i State Department of Education, RS 94-5555.
- Peter, L., & Hirata-Edds, T. (2006). Using Assessment to inform instruction in Cherokee language revitalization. *International Journal of Bilingual Education and Bilingualism*, 9(5), 643-658.
- Peter, L., Sly, G., & Hirata-Edds, T. (2008, October). *Constructing language assessments in indigenous languages to inform instruction*. Power point presentation at the CARLA Conference, Immersion Education: Pathways to Bilingualism and Beyond, Saint Paul, Minnesota.
- Raymond, M. (2008, October). *The Cherokee nation and its language: tsalagi ayeli ale uniwonishisdi*. Paper presented at the CARLA Conference, Immersion Education: Pathways to Bilingualism and Beyond, Saint Paul, Minnesota.
- Slaughter, H. B., & Watson-Gegeo, K. (1988). *Evaluation report for the first year of the Hawaiian language immersion program: A report to the Planning and Evaluation Branch, State of Hawai'i, Department of Education*. Honolulu, Hawai'i: University of Hawai'i, Mānoa.
- Slaughter, H. B., Warner, S. N., & Palmeira, W. K. (1989). *Evaluation report for the second year of the Hawaiian language immersion program: A report to the Planning and Evaluation Branch, State of Hawai'i, Department of Education*. Honolulu, Hawai'i: University of Hawai'i, Mānoa.
- Slaughter, H. B., Lai, M. K., Warner, S. N., & Palmeira, W. K. (1990). *Evaluation report for the third year, 1989-1990, of the Hawaiian language immersion program: A report to the Planning and Evaluation Branch, State of Hawai'i, Department of Education*. Honolulu, Hawai'i: University of Hawai'i, Mānoa.
- Staton, R. (2005, March 14). Language revival: Hawaiian rates as the nation's only growing indigenous tongue. *Honolulu Star Bulletin*. Retrieved October 16, 2009 from <http://starbulletin.com/2005/03/14/news/story7.html>.
- Thompson, L., Boyson, B., & Rhodes, N. (2006). *Administrator's Manual for CAL foreign language assessments, grades K-8*. Washington DC: Center for Applied Linguistics & Iowa State University National K-12 Foreign Language Resource Center.



Warner, S. (1996). I ola ka 'ōlelo i nā keiki: Ka 'apo 'ia 'ana o ka 'ōlelo Hawai'i e nā keiki ma ke kula kaiapuni. *Dissertation Abstracts International*, (UMI No. 9629864)

## Appendix A: Hawaiian Oral Language Proficiency Assessment Rubric (‘Ōlelo Hawai‘i)

### PAKUHI ANALOI MĀKAU ‘ŌLELO WAHA NO KA ‘ŌLELO HAWAI‘I

Pili i ka Mana‘o			Pili i ke Kani		Kuana‘ike Hawai‘i		
Mākau Ho‘oka‘a‘ike	Hua‘ōlelo	Pilina‘ōlelo	Puana	Poeko	Kūpa‘a ‘Ōlelo	Ke ‘Ano Hawai‘i o ka ‘Ōlelo Hawai‘i ‘ana	
<b>3</b>	<ol style="list-style-type: none"> <li>Ho‘opuka ‘ia nā mana‘o ma nā hopuna‘ōlelo piha.</li> <li>Mōakāka nō nā mana‘o ke lohe aku.</li> <li>Aia ma kāna pane ‘ae‘oia ka nui o nā ‘āpana o ia māhele o ka hō‘ike. (ho‘olauna, pūka‘ina ki‘i)</li> <li>Hiki ke hana nona iho, ‘a‘ohe kauka‘i i ka mea loiloi ma ka ho‘oholo ‘ana i ke kama‘ilio ‘ana.</li> </ol>	<ol style="list-style-type: none"> <li>Lawa ka ‘ikena hua‘ōlelo e ho‘okō pono ai i ka hana hō‘ike. Ho‘ohana pū ‘ia nā hua‘ōlelo o nā pō‘aiapili ma ‘ō aku.</li> <li>Ho‘ohana pololei mau ‘ia nā hua‘ōlelo.</li> <li>Kāka‘ikahi (‘a‘ohe paha) ka wā e pono ai e kū a no‘ono‘o ma ke koho ‘ana i ka hua‘ōlelo.</li> </ol>	<p>Ho‘ohana ‘ia nā pilina‘ōlelo like ‘ole he nui.</p> <p>Kāka‘ikahi ka wā e ho‘ohana pololei ‘ole ‘ia ai ka pilina‘ōlelo.</p>	<p>Puana pono mau ‘ia nā ‘āpana hakalama, nā huēwoela, ka ‘okina, ke kahakō, nā kāpana, a me nā hualeo.</p>	<p>Lohe ma‘amau ‘ia nā hi‘ohi‘ona poeko Hawai‘i:</p> <ul style="list-style-type: none"> <li>ka wali</li> <li>ke ki‘ina leo</li> <li>ke kālele</li> <li>ka pana (wikiwiki/mālie)</li> <li>ke kū ‘ana i ka wā kūpono.</li> <li>‘a‘ole ho‘ohana ‘ia ka ‘ōlelo “um”</li> </ul>	Kūpa‘a mau i ka ‘ōlelo Hawai‘i	<p>Lohe ‘ia 2 a ‘oi hi‘ohi‘ona o kēia analoi:</p> <ol style="list-style-type: none"> <li>ka no‘ono‘o Hawai‘i/‘ike ku‘una (“kau” ma luna o ke ka‘a, “holoholo” ma kahi o “lawai‘a”)</li> <li>loina/meiwi (kūpina‘i, ‘ēko‘a, helu, welina)</li> <li>‘ikeoma, ‘ōlelo kaulana, ‘ōlelo no‘eau</li> <li>kuana‘ike ‘ohana</li> </ol>
<b>2</b>	<ol style="list-style-type: none"> <li>Ho‘opuka ‘ia nā mana‘o ma ka pae hopuna‘ōlelo a māmalala‘ōlelo paha i ka hapanui o ka manawa.</li> <li>Mōakāka ka nui o nā mana‘o ke lohe aku.</li> <li>Aia ma kāna pane ‘ae‘oia kekahi hapa o nā ‘āpana o ia māhele.</li> <li>Kauka‘i iki ‘ia ka mea loiloi ma ka ho‘oholo ‘ana i ke kama‘ilio ‘ana.</li> </ol>	<ol style="list-style-type: none"> <li>Lawa ka ‘ikena hua‘ōlelo e ho‘okō ai i ka hana hō‘ike.</li> <li>Ho‘ohana pololei ‘ole ‘ia nā hua‘ōlelo i kekahi manawa.</li> <li>Kū i kekahi manawa e no‘ono‘o i ka hua‘ōlelo a loa‘a iā ia nona iho i ka hapanui o ka manawa.</li> </ol>	<p>Ho‘ohana ‘ia nā pilina-‘ōlelo like ‘ole.</p> <p>Lohe ‘ia kekahi mau hemahema pilina‘ōlelo.</p>	<p>Hemahema ka puana i ia mau hi‘ohi‘ona ‘ōlelo o luna a‘e nei i kekahi manawa.</p>	<p>Lohe ‘ia nā hi‘ohi‘ona poeko Hawai‘i (i helu ‘ia ma luna a‘e nei) a lohe pū ‘ia ka hi‘ohi‘ona ‘ōlelo ‘ē, e like ho‘i me “um”.</p>	Kūpa‘a ma ka ‘ōlelo Hawai‘i i ka hapanui o ka manawa. Komo na‘e ka ‘ōlelo ‘ē i kekahi manawa.	<p>Lohe ‘ia 1 la‘ana o nā ‘ano Hawai‘i o ke kama‘ilio ‘ana i helu ‘ia ma luna a‘e nei.</p>
<b>1</b>	<ol style="list-style-type: none"> <li>Mumuku ka ‘ōlelo i ho‘opuka ‘ia ma nā hopuna‘ōlelo piha ‘ole a me nā hua‘ōlelo a māmalala‘ōlelo paha.</li> <li>Mōakāka ‘ole nā mana‘o ke lohe aku.</li> <li>Nele nō ka pane ‘ae‘oia i ka nui o nā ‘āpana o ia māhele o ka hō‘ike.</li> <li>Kauka‘i nui ‘ia ka mea loiloi ma ka ho‘oholo ‘ana i ke kama‘ilio ‘ana.</li> </ol>	<ol style="list-style-type: none"> <li>‘A‘ole i lawa nā hua‘ōlelo e ho‘okō ai i ka hana hō‘ike.</li> <li>Ho‘ohana pinepine ‘ia nā hua‘ōlelo pololei ‘ole no ka pō‘aiapili.</li> <li>Kū pinepine e no‘ono‘o i ka hua‘ōlelo. ‘O ka pono o ke kōkua ka hopena i kekahi manawa.</li> </ol>	<p>‘A‘ole i nui nā pilina-‘ōlelo like‘ole i ho‘opuka ‘ia.</p> <p>Nui nō nā hemahema pilina‘ōlelo.</p>	<p>Pinepine a ahuwale nā hemahema puana o ia mau hi‘ohi‘ona ‘ōlelo o luna a‘e ala.</p>	<ul style="list-style-type: none"> <li>Kāka‘ikahi nā hi‘ohi‘ona poeko Hawai‘i i helu ‘ia ma luna a‘e ala.</li> <li>Lohe nui ‘ia ka hi‘ohi‘ona ‘ōlelo ‘ē.</li> <li>He hana nui ka ho‘opuka mana‘o ‘ana: he ‘ā‘ā ka leo.</li> </ul>	Komo mau nā hua‘ōlelo ‘ōlelo ‘ē ma ka ‘ōlelo.	<p>‘A‘ole lohe ‘ia ke ‘ano Hawai‘i o ke kama‘ilio ‘ana i wehewehe ‘ia ma luna a‘e ala.</p>

## Appendix B: Hawaiian Oral Language Proficiency Assessment Rubric (English Version)

### HAWAIIAN ORAL LANGUAGE PROFICIENCY ASSESSMENT RUBRIC

Semantics			Phonology		Kuana'ike / Worldview		
Communicative Skills	Vocabulary	Grammar	Pronunciation	Fluency	Language Steadfastness	Cultural and Linguistic Authenticity	
<b>3</b>	<ul style="list-style-type: none"> <li>Speaks in complete sentences.</li> <li>Ideas expressed are easily grasped by the listener.</li> <li>Student's independent response includes most aspects of the assessment task (introduction, picture series).</li> <li>Is able to independently direct his/her own communications. Speaks without relying on prompts or assistance.</li> </ul>	<ul style="list-style-type: none"> <li>Word knowledge encompasses contexts of speech, which fulfill and go beyond the task.</li> <li>Vocabulary is consistently used correctly.</li> <li>Stops to search for words are rare if occurring at all.</li> </ul>	<ul style="list-style-type: none"> <li>Uses a wide variety of grammar patterns.</li> <li>Grammatical errors, if any, are infrequent.</li> </ul>	<ul style="list-style-type: none"> <li>Consistently pronounces hakalama, vowel blends, 'okina, kahakō, and phonemes correctly.</li> </ul>	<ul style="list-style-type: none"> <li>Speech embodies features of conversational Hawaiian language fluency: ease/comfort of speaking proper inflection, rhythm (speed and slowness as appropriate), proper pauses, absence of "um".</li> </ul>	<ul style="list-style-type: none"> <li>Steadfast adherence to Hawaiian language.</li> </ul>	<ul style="list-style-type: none"> <li>2 or more unique Hawaiian language features are present:               <ol style="list-style-type: none"> <li>Hawaiian thought/traditional knowledge ("to get on the car"; "holoholo" in place of "going fishing"; etc)</li> <li>Language traditions/oratorical features (recites mo'okū'auhau, complementary pairs, opposites, etc...)</li> <li>Idioms, famous sayings, proverbs</li> <li>Use of traditional family terms correctly</li> </ol> </li> </ul>
<b>2</b>	<ul style="list-style-type: none"> <li>Speaks mostly at the phrase or sentence level.</li> <li>Ideas expressed are mostly clear in meaning.</li> <li>Some aspects of the task are included in the student's independent response.</li> <li>The student requires some assistance in producing responses to the test items.</li> </ul>	<ul style="list-style-type: none"> <li>Word knowledge is limited to the task at hand.</li> <li>Vocabulary is not always used correctly.</li> <li>Stops to search for words are sometimes necessary and usually result in finding a workable word choice.</li> </ul>	<ul style="list-style-type: none"> <li>Uses a variety of grammar patterns.</li> <li>Grammar errors are present.</li> </ul>	<ul style="list-style-type: none"> <li>Mispronounces some aspects of hakalama, vowel blends, 'okina, kahakō, syllables, and phonemes.</li> </ul>	<ul style="list-style-type: none"> <li>Features of conversational Hawaiian language fluency (listed above in row 3) are present along with some interference from foreign language features, e.g., "um".</li> </ul>	<ul style="list-style-type: none"> <li>Exhibits Hawaiian language steadfastness with occasional foreign language intrusion.</li> </ul>	<ul style="list-style-type: none"> <li>Inconsistently exhibits the above features of speaking Hawaiian in a Hawaiian way (1 instance).</li> </ul>
<b>1</b>	<ul style="list-style-type: none"> <li>Speaks in brief and incomplete sentences, sometimes involving only one word or phrase.</li> <li>Ideas expressed are unclear.</li> <li>Few aspects of the task are included in the student's independent response.</li> <li>Assistance and prompts are often required.</li> </ul>	<ul style="list-style-type: none"> <li>Word knowledge is not sufficient to adequately fulfill the task.</li> <li>Vocabulary is often used incorrectly.</li> <li>Stops to search for words are frequent and necessary, and sometimes lead to the need for assistance.</li> </ul>	<ul style="list-style-type: none"> <li>Uses a limited number of grammar patterns.</li> <li>Grammatical errors are frequent.</li> </ul>	<ul style="list-style-type: none"> <li>Pronunciation errors are frequent and obvious.</li> </ul>	<ul style="list-style-type: none"> <li>Features of conversational Hawaiian language fluency (listed above in row 3) are noticeably lacking.</li> <li>Foreign language fluency features frequently interfere.</li> <li>Student speaks haltingly ('ā'ā ka leo).</li> </ul>	<ul style="list-style-type: none"> <li>Frequently interjects foreign language vocabulary.</li> </ul>	<ul style="list-style-type: none"> <li>Does not exhibit the Hawaiian language features listed in row 3 above.</li> </ul>