USING PERFORMANCE ASSESSMENTS TO INCREASE MOTIVATION IN THE FOREIGN LANGUAGE CLASSROOM AT THE HIGH SCHOOL LEVEL

Doris M. Waletzko

A Research Project

Approved:

Suzanne C Griffith — 11/11/2016
Research Project Advisor — Date

Suzanne C Griffith — 11/22/2016
Educational Leadership Department Chair — Date

Dean G — 11/16/16
Dean of Faculties and Graduate Studies — Date

USING PERFORMANCE ASSESSMENTS TO INCREASE MOTIVATION IN THE FOREIGN LANGUAGE CLASSROOM AT THE HIGH SCHOOL LEVEL

A Research Project Report

Presented to

the Graduate Faculty at the

University of Wisconsin – Superior

In Partial Fulfillment

of the Requirement for the Degree

Master of Science in Education – Instruction

August 2016

by

Doris Dioses Waletzko

Abstract

This work describes the importance of Integrated Performance Assessments in the foreign language field can promote changes in the learning and teaching of a foreign language. But new practices can be slow to implement due to lack of distribution of knowledge, time and budget constraints, and teachers' attitudes and practices in the classroom. The national foreign language standards were created as a guideline for schools as to the knowledge students should acquire. Integrated Performance Assessments (IPA) were developed to connect this goal to the classroom as a process to evaluate students' learning. There is a need to clarify the use of IPA compared to traditional end of chapter tests and to what extent these tests serve to promote the standards. The researcher compared two approaches for assessing foreign language learning through comparison of the students' grades from IPA to the end of chapter test grades from a convenience sample of students using grades recorded in a teacher's grade book. There was wide deviation in traditional end of chapter scores which could indicate a problem in evaluating students' knowledge using this more traditional scoring method. The students as a whole received better grades on the IPA as there was a narrow deviation which appeared to show that students had more success in their learning according to the IPA. Use of IPA may have more correctly reflected students' foreign language learning and proficiency.

Acknowledgements

As a professional of education, I would like to acknowledge the many opportunities that I have encountered in my teaching career which have allowed me to grow as a professional in education. Those opportunities, at times challenging, have taken me to various settings from kindergarten to college level, from schools in South America, to schools in USA. In an attempt to weigh the pros and cons of the exposure to various school settings, without a doubt I have to say that the good – and the not so good – experiences have played an important role in my formation and transformation as a teacher. Thanks to those opportunities, I have met wonderful people – teachers, parents, students, and administrators – who have helped me grow and supported me not only as a teacher but as a person and professional who strongly advocates for second language learning.

Thank you to Lisa Hendrickson and Karen Luond Fowdy, world language consultants, who through their Teaching What Matters workshop inspired the much needed change to the teaching and assessing student progress in our area as we face the challenges of the 21st century. I would also like to thank my friend Kathy who has been a strong support since I moved to the United States and wanted to continue my teaching career that I started in Lima, Peru. The paperwork, tests, classes, job applications, my limitations in the second language, were at times overwhelming but she has always been there willing to give me a hand, willing to help, and to listen to the many amazing stories a teacher has to share. I am a lucky person who has not only found an inspiring profession but someone who has found true friendship. Last, but not least, thank you to my family, my bother in law Jim proof reading my work providing a related teaching story for every chapter, thank you to my husband Scott for your moral and technical support, and to all the first generation of professional women in my family for your perseverance.

TABLE OF CONTENTS

	Page
Abstract	ii
Acknowledgments	iii
Table of Contents	iv
List of Tables and Figures	vi
CHAPTER 1 – THE PROBLEM	
Introduction	1
Statement of the Problem	2
Research Question	3
Importance of Study	3
Purpose Statement	4
Assumptions	4
Limitations	5
Delimitations	5
Definitions of Terms	6
Conclusion	9
CHAPTER 2 –LITERATURE REVIEW	
Introduction	11
National Standards	12
Integrated Performance Assessment	19
Teachers' Perspectives	31
Conclusion of Literature Review	37

CHAPTER 3 – RESEARCH DESIGN

Introduction 40

Setting and Participants 41

Instrumentation 41

Procedure 44

Data Collection and Analysis 45

CHAPTER 4 – RESULTS

Analysis of Data 47

Conclusion of Results 50

CHAPTER 5 – DISCUSSION

Findings and Interpretations 51

Limitations 55

Recommendations for Further Research 56

Conclusion 56

REFERENCES 58

APPENDICES

Appendix A – World-Readiness Standards for Language Learning 64

Appendix B – ACTFL Can Do Statements for Interpersonal Communication 65

Appendix C – Fairfax PALS Rubric 66

Appendix D – Grammar Test 67

Appendix E – *t*-Scores Data 70

LIST OF TABLES AND FIGURES

Table 1 48

Table 2 49

Table 3 49

Figure 1 50

Figure 2 52

Figure 3 53

CHAPTER 1

INTRODUCTION

After a few years in the educational arena, educators may wonder if their practices and their department's practices are supported by research or wonder what the research has to say about present practices. With the advent of new initiatives in the field of foreign languages learning in the arena of assessment there is a need in many school districts for clarification as to how best to put together new practices and find value in the process. Within the field of foreign languages learning these changes raise questions: Where is the profession going with instruction; what methods are most effective; what data is there to support new instructional approaches; and what are the best practices? Integrated Performance Assessment (IPA) falls under the umbrella of 'thematic focus and backward design using comprehensible input' and could be one of the newer and more effective answers as stated by Sandrock (2010) and VanPatten (2015).

According to Moeller, Creswell, and Saville (2016) the past 50 years in language study have brought a transition from testing of grammar and vocabulary to an integration of original material from the studied language to a focus on communication. Sandrock (2010) stresses the importance of the performance assessment practice for both students and teachers. Students are prompted to use the language they are learning and teachers enhance their instruction with effective core practices resulting in evidence of learning as the ultimate goal. Vélez-Rendón (2002) states "While it is true that many second language teacher educators are seasoned and reflective thinkers, it is also true that many need to re-think their roles and renew their practices" (p. 464).

Foreign language departments across the globe have adopted changes in their curriculum and in the evaluation of speaking, listening, writing, and reading. For example, collaboration is

taking place between the American Council on the Teaching of Foreign Languages (ACTFL) and the Common European Framework of References for Languages (CEFR) on mixed methods research in the field of language assessment, as described by Moeller et al. (2016). In the United States, the ACTFL has been fundamental in the development of many key initiatives in language learning. The ACTFL Proficiency Guidelines (PGs) (ACTFL, 2012a), Can-Do statements (2012b), and Performance Guidelines for K-12 Learners (ACTFL, 1998) are used in conjunction with the National Standards for Foreign Language Education Project (NSFLEP) to improve student language usage and cultural understanding (ACTFL, 2012a). Some school districts have incorporated these new ways to assess students while others continue using traditional practices including end of chapter testing to assess language learning where proficiency may not be the ultimate goal for the program.

Statement of the Problem

New knowledge and information promote changes in the learning and teaching of a foreign language but these changes arrive at different speeds to different school districts. In 1986 the ACTFL adopted PGs for foreign language learning assessment. In 1997 the Wisconsin Department of Public Instruction (WI DPI) (Keen, Grady, & Sandrock, 2001) recommended following these guidelines but did not require school districts to do so. In the school district in northern Wisconsin, where this study took place, language teachers began to adopt the idea of PGs at the high school level for the school year 2014-2015, almost three decades after ACTFL was released and 17 years after they were recommended by the WI DPI (Keen et al., 2001). Van Patten (2015) targeted a similar issue in his paper where he stated that advances in language acquisition science and language teaching were slow to be implemented in education and in some cases were not recognized at all.

In recent years, foreign language departments have adopted the ACTFL's approach to the evaluation process plus a theme-based curriculum with primary goals of communication, cultural understanding, career opportunities, and improvement of language skills based on using primarily authentic material. The Wisconsin standards for foreign language cite similar goals (Keen et al., 2001). While this approach has evidence that showed it to be effective (Adair-Hauck, Glisan, Koda, Swender, & Sandrock, 2006), there is a local need to clarify the purpose of traditional end of chapter tests that continue to be used and to what extent these serve to promote these goals and standards.

Research Question

This thesis examined if using IPA would result in measuring greater learning among students and higher scores on evaluation of their foreign language skills when compared to the results a chapter knowledge test would derive. The study asked how the use of IPA compared to an end of chapter test on high school students' success in foreign language learning as reflected in students' grades.

Importance of Study

While it has been shown that IPA can be an effective way to evaluate students there has been a lack of continuing research in this area (Kissau & Adams, 2016). This study is important because it is additional research on IPA at the high school level, especially for districts at the early stages of adopting performance assessments according to the national standards for learning languages. The present study was a district specific study to compare IPA to the traditional testing that continued to be used in the district's high school foreign language classes. It examined the effect on students through their grades depending upon whether IPA or end of chapter was the type of evaluation being used.

Purpose Statement

The main purpose of the study was assess high school students' foreign language communication and learning in a school district in northern Wisconsin. It compared two approaches for assessing foreign language learning through comparison of the students' grades from IPA to those of the end of chapter test grades. In the second year of implementation of IPA in the school district there was a need to clarify if the continued use of one or both of these assessment measures served to promote the district's goals and standards and allowed the students to achieve the grade level that represents their skills in the language.

Data was gathered from the teacher's grade book of five classes, a total of 147 secondary students in grades 9 through 12, to examine if students in the second year of Spanish foreign language learning were successful when learning about the chapter topic (shopping). This class was instructed using backward design for the shopping unit for six weeks. At the beginning of the unit the instructor identified what students would be expected to do at the end of the unit. These expectations were based on the minimal standards for foreign language learning from the ACTFL Performance Guidelines for K-12 learners, IPA rubrics, and Can-Do statements. The researcher hypothesized that the students would have better outcomes and more accurate evaluation of their learning in accordance with the national and state standards in the IPA results when compared to traditional end of chapter testing results.

Assumptions

Throughout the execution of this study, it was assumed that students were going to do their best to participate honestly in two types of testing, traditional end of chapter test and oral performance assessment. The tests were presumably unbiased and meant to collect relevant data.

For all assessments the anonymity of the participants was protected. It was also assumed that the learning and teaching processes covered the foreign language standards as was stated.

Limitations

This study was limited to a convenience sample of five sections of high school students, with the same teacher, enrolled in the second level of a foreign language class with results from a five week unit. An average grade of C or better in Math and English classes was required to enroll for the first year in a language class. Subsequently, an average grade of C or better in the first level was required to advance to the second level and so on.

Another limitation to this study was that only one school district from northern Wisconsin was included. This school was mainly comprised of students identified as Caucasian (92%) with 8% from other groups (Black 2%, Hispanic 1%, Asian 2%, and Native American Indian 3%), according to the Public School Review website http://www.publicschoolreview.com with reference to the 2014 National Center for Education Statistics, WI Dept. of Education source. These percentages were not necessarily reflected in the specific classes in this study. Since the minority enrollment of 8% of the student body (majority American Indian) was less than the state average of 28%, the findings of this study are limited because the enrollment did not reflect national or state diversity.

Delimitations

There are several delimitations that were present in this study. This study did not account for socio-economic status, gender, or race within the student population. This study did not intend to evaluate the teacher's performance or level of language discourse. The sampling procedure used by the researcher was a convenience sampling. The study does not differentiate

among the five sections of level two foreign language learning as it considered the whole group as one. Future studies should include more foreign language instructors as it was a limitation that all five class sections were instructed by same teacher which may have influenced the findings.

Definitions of Terms

The following definitions will be useful in understanding the review of the literature and the study:

Authentic Assessment: Assessment that mirrors the tasks and challenges faced by individuals in the real world (Glisan, Adair-Hauck, Koda, Sandrock, & Swender, 2003, p. 37).

Authentic Materials/Texts: Oral or printed materials/texts that have been produced by and for native speakers of the target culture for non-instructional purposes; e.g., newspapers, magazines, books, television programs, radio broadcasts (Glisan et al., 2003, p. 37).

Backward Design: The teacher starts with desired outcomes and acceptable evidence and then plans appropriate learning activities to reach those outcomes (Wiggins & McTighe, 2005).

Can-Do Statements: "…describe the specific language tasks that learners are likely to perform at various levels of proficiency. Can-Do Statements mirror much of what is addressed in the Common Core State Standards (CCSS) for English Language Arts (ELA) and Literacy. Learning another language…, involves learning to read, write, speak, listen, comprehend, communicate, and use conventions of language" (ACTFL, 2012b, p. 2).

Foreign Language: A language is considered foreign if it is learned largely in the classroom and is not spoken in the society where the teaching occurs. Study of another language allows the individual to communicate effectively and creatively and to participate in real-life situations through the language of the authentic culture itself. Learning another language provides access into a perspective other than one's own, increases the ability to see connections across content areas,

and promotes an interdisciplinary perspective while gaining intercultural understandings (Moeller & Catalano, 2015, p. 237).

Foreign Language Learning and Teaching: refer to the teaching or learning of a nonnative language outside of the environment where it is commonly spoken. A distinction is often made between 'foreign' and 'second' language learning (Moeller & Catalano, 2015, p. 327).

Formative Assessment: "Learning checks, guided activities applications of skill and knowledge, with more teacher assistance, intervention and support; designed to help shape learners' understanding and skills while there is an opportunity for the teacher and learners to work together to bring about further development and improvement." (Sandrock, 2010, p. 13).

Integrated Performance Assessments: "…ask students to use languages for real purposes: sharing new information, exchanging opinions, presenting ideas to a specific audience, preparing a letter of application or introduction, understanding the point of view of a speaker or author and comparing it to one's own, or skimming a website to find needed information. A well designed performance assessment task will generate these genuine acts of communication". (Sandrock, 2010, p. 2)

Interpersonal Tasks: "Interpersonal tasks are two-way, interactive activities, such as face to face or telephone conversations and spontaneous written correspondence, such as emails or text messaging. In oral interpersonal communication, speakers communicate in a spontaneous manner and do not use a written script" (Sandrock, 2010, p. 8).

Interpretive Communication: A reader or listener is engaged in understanding the meaning of oral, written, or other cultural texts (i.e., film, radio, television, newspapers, magazines, or literature) when the author of these texts is not present and meaning cannot be negotiating (Glisan, et al., 2003, p. 38).

Mixed Methods Research: "….in the field of language assessments involves the collection and combination of quantitative (numbers oriented) data, and qualitative (text and stories oriented) data, in the test development and assessment process. …Mixed methods is not simply formative and summative evaluation. It has more components, such as mixed methods questions, mixed methods designs, and other features that make it a systematic methodology for research". (Moeller et al., 2016, p. 36).

Mode: A manner of communicating; the National Standards specify three modes of communication: interpersonal, interpretive and presentational (Clementi & Terrill, 2013, p. 114).

Performance: is the ability to use language that has been learned and practiced in an instructional setting. Coached by an instructor, whether in a classroom or online, or guided by instructional materials, performance refers to language ability that has been practiced and is within familiar contexts and content areas. The practice and assessment of performance should reflect authentic, real world use of language, even though the language is learned and practiced in some type of learning environment (ACTFL, 2012c, p. 4).

Presentational Communication: Individuals engage in one way oral or written communication that presents information to an audience for interpretation with no possibility of negotiating meaning (Glisan et al., 2003, p. 38).

Proficiency: What individuals can do with the language in terms of speaking, writing, listening, and reading in real world situations in a spontaneous, non-rehearsed context (ACTFL, 2012c, p. 4).

Rubric: Written and shared criteria for judging performance that indicate the qualities by which levels of performance can be differentiated, and that anchor judgments about the degree of success on a student assessment (Glisan et al., 2003, p. 39).

Second Language: A second language implies that the learner resides in an environment where the acquired language is spoken. In the area of research, the term second language acquisition (SLA) is a general term that embraces foreign language learning and investigates the human capacity to learn languages other than the first language once it has been acquired (Moeller & Catalano, 2015, p. 327).

Second Language Acquisition: All non-native language acquisition.

Summative Assessment: "End-of-unit or end-of-course assessment of language performance; a demonstration of what students should be able to do on their own as a result of the unit of instruction" (Sandrock, 2010, p. 13).

Target Language: The language other than one's native language that is being learned (Clementi & Terrill, 2013 p. 114).

Theme: A unifying subject or idea of an instructional unit (Clementi & Terrill, 2013, p. 114).

Traditional/End of Chapter/Grammar Test: Discrete point test of grammar and vocabulary.

Washback Effect: The impact that tests have in teaching and learning. One of the goals of seminal study was to demonstrate the effect of performance assessments on teachers' perceptions regarding their instructional practice (Adair-Hauck et al., 2006, p. 360).

World-Readiness Standards for Learning Languages: defines the central role of world languages in the learning career of every student. The five goal areas of the Standards establish an inextricable link between communication and culture, which is applied in making connections and comparisons and in using this competence to be part of local and global communities (ACTFL, 2015).

Conclusion

The purpose of this research was to gather data on two assessment methods for foreign language proficiency in a school district in northern Wisconsin. It compared two approaches for

assessing foreign language learning through comparison of the students' grades from IPA to the end of chapter test grades. In the second year of implementation of the IPA in the school district there was a need to clarify if the continued use of one or both of these assessment measures served to promote the district's goals and standards and to allow the students to achieve the grade level that represents their skills in the language.

CHAPTER 2

LITERATURE REVIEW

In a globally connected world, learning a second language is undoubtedly an important and a skill to have. The opportunities to enjoy new places, understand cultures, be more marketable for a job, and be prepared for upper level education count among the many benefits of learning a second language. Even though the advantages are vast, only 18.5% of the United States public school population was enrolled in foreign language class in 2007-2008 (Committee for Economic Development, n.d; Partnership for 21st Century Skills [P21], 2011). Consequently, this data reflects a huge gap between what is happening in our schools and what the U.S. government and the world market are apparently demanding. Solid language instruction is needed to prepare global citizens who can communicate in a language other than English. Even though 18.5% is not a high percentage of the entire school population, it is still important that educational efforts should aligned the best practices for those students to be able to achieve all the standards for foreign languages.

This study proposed to examine students' success and achievement in foreign language learning, in particular with regard to Interpersonal Communication Standard 1.1 that required students to engage in conversations, to provide and obtain information, to express feelings and emotions, and to exchange opinions. It compared IPA to traditional end of the chapter testing. To that end, the articles presented are a review of relevant research.

Three main areas are addressed in the review of the literature. The first section describes research related to national standards and core practices in language teaching. The second section of the literature review includes research that studied the potential benefits of language performance assessment taking into account the goal to achieve proficiency in meeting the

standards for foreign language in concrete ways. The third section explores teacher practices and attitudes towards incorporating advances in language research and education into their classrooms.

National Standards

The National Standards for Foreign Language Education was part of a mandate by the U.S. Department of Education to set standards for all core subjects. The national standards for foreign language learning were developed in 1996 through a collaborative effort of 10 foreign language associations with a basic philosophy that foreign language learning in the United States was a necessity. "Language and communication are at the heart of human experience. The United States must educate students who are equipped linguistically and culturally to communicate successfully in a pluralistic society and abroad. This imperative envisions a future in which ALL students will develop and maintain proficiency in English and at least one other language, modern or classical" (National Standards for Foreign Language Education Project [NSFLEP], 2006, p. 7).

What are the effects of the national standards on teaching and learning, on state and local curriculum, on classroom instruction, on methods courses, and on professional development since its conception? To answer this question, ACTFL was given a grant in 2008 by the federal government. The grant funded the project for a three-year period in which, during the first year, three task forces would conduct literature reviews, develop a survey, and survey members and organizations. The tasks for the second year would be to analyze data from the survey and to interpret that data according to literature search. In the last year, the task forces would draw implications from the data results and literature search to pursue directions in the future (Abbott & Phillips, 2011).

In year one of this study, according to the NSFLEP task forces, the survey was electronically delivered to 2,134 respondents who identified as language teachers, methods instructors, district supervisors/program chairs, or state supervisors (ACTFL, 2011). Of those respondents, 89% identified as language teachers. The respondents who said they had no familiarity with the standards were not asked additional questions (ACTFL, 2011).

The NSFLEP task forces detailed the results of the survey finding that the total respondents were 72% female, that 67% of the teachers had 11 or more years of teaching experience, that 69% taught Spanish, that 62% taught high school, and that 89% were familiar with the national standards (ACTFL, 2011). The NSFLEP survey asked respondents about adoption of the standards and found 78% were aware that the standards of their state were based on the national standards and 80% said their local or departmental standards were based on the national standards. There were 75% or more of the teachers that agreed that the national standards had influenced different aspects of their teaching, according to ACTFL (2011).

The NSLEP survey (ACTFL, 2012) also questionned administrators and reported that 58% of administrators replied that they used performance based assessment and teaching of 21st century skills. State supervisors replied that their state emphasized the Communication standard and to some extent the Culture standard but felt there was not enough time for actual classroom implementation of all the standards. ACTFL (2011) found that many respondents indicated that there were too many standards and had confusion between goals and standards.

While district supervisors said their districts had reformed curriculum with the standards in mind, ACTFL (2011) noted that reform in assessments had not been accomplished. It reported that in 43% of districts, assessments were said to align with standards and 38% said they were partially aligned. Only about half of the teachers surveyed had received professional

development (PD) on the standards, and most of this PD was on the Communication standard and the second highest PD was on Culture. The researchers thought that may have been because those two standards most aligned with traditional teaching and because of that there were more experts in those fields. ACTFL (2011) indicated, that according to teachers, the PD gave them more realization of the interpretive part of the Communications standard. Most of the teachers reported that they had implemented something new that was related to the standards into their classrooms after PD but few continued to reflect on their teaching or pursued further research. Furthermore, sixty percent of teachers said there was no follow up support by the administration after PD; only 15% reported being given time for follow up work.

In their analysis of the survey data from the NSFLEP task forces, Abbott and Phillips (2011) stated that when district supervisors answered the question about limitations on planning PD that the main limitations were time (68%) and budget (67%). Although the question was not addressed specifically in the survey, 11% stated in the open-ended question that there were teachers that did not want to change. Abbott and Phillips also found that administrators responded that, in addition to time and budget, one of their greatest challenges overall in implementing new standards and curriculum were the teachers that wanted to continue to rely on the textbook, use discrete point testing of grammar, and not work collaboratively. One of the questions the researchers said should be asked in the future is if there is a common thread in teachers that do not want change. Challenges with budget constraints were partly because foreign language has not been considered a core area and because the No Child Left Behind Act shifted priorities away from foreign language offerings, according to Abbott and Phillips. However, the authors did report that district supervisors had their greatest successes with teachers working collaboratively to write their own curriculum; making PD available with outside consultants,

workshops, and new material; and aligning assessments with teaching. Abbott and Phillips recapped that communication is the goal in IPAs but content and culture can also be included so that an IPA incorporates 3 of the 5 Cs.

Another study that surveyed teachers and administrators across the country to inquire on the state of foreign language education in the schools was done by Pufahl and Rhodes (2011) in conjunction with the Center for Applied Linguistics. The Center for Applied Linguistics (CAL) was also concerned with students having the ability to participate in a global culture and started with a national K-12 survey in 1987 to be done every ten years. CAL sent another questionnaire to 5000 public and private schools in the country in 2008 (Pufahl & Rhodes, 2011).

According to Pufahl and Rhodes (2011), the questionnaire was created with two versions, one for elementary and one for secondary schools, with both online and paper copies. The questions were formatted to be similar to the questionnaires that had been used on the previous two surveys in order to maintain consistency over time in what was surveyed. The researchers tested the questionnaire before the larger survey with teachers and principals; the online survey was tested for usability also.

Pufahl and Rhodes (2011) used a stratified random sample analyzed by type of school: public or private, elementary or secondary. Additional strata were the location of schools (metro status and geographic area) and distribution of student populations pertaining to socioeconomic status and percent of minorities in the school. For the survey, 76% of the schools responded. In the final results, Pufahl and Rhodes reported that 2,668 elementary schools and 1,002 secondary schools participated and the schools were distributed approximately one third each from rural, suburban, and urban areas, with approximately two thirds public and one third private.

Pufahl and Rhodes (2011) examined the data for 2008 and some elements were compared statistically to the survey from 1997 with a *t*-test that looked for significant changes over the past decade. Pufahl and Rhodes found that fewer public elementary schools offered foreign language for students than ten years earlier; dropping from 31% to 25%. Middle schools dropped from 75% of schools to 58% of all schools. The authors found there was little change in the number of high schools who offered foreign language instruction. Few public elementary schools planned to offer foreign language in the near future, although 25% of private schools planned to begin to offer foreign language learning. Pufahl and Rhodes noted that respondents gave reasons not to offer foreign languages at the elementary level which included not being considered a core subject and no funding by the district. It was also mentioned that there was difficulty finding qualified language teachers. In general, Pufahl and Rhodes found there was more interest in offering language learning in private schools than in public schools; at the high school level, 73% of students in private schools and 48% in public schools were enrolled in the foreign language programs.

Pufahl and Rhodes (2011) also examined what languages were offered and type of curriculum. Schools at all levels had an increase in use of authentic materials from the studied language, including increased technology usage. They found little continuity of instruction from elementary to middle school to high school. For example, students in immersion in elementary would find only level one offered at their middle school. Therefore with students beginning at different levels, students could find that their fellow students in a class had more or less experience in the language. In spite of the greater percentage of private schools that offered foreign language Pufahl and Rhodes reported there were fewer teachers in private schools that were certified to teach in the language and fewer who had continued training or PD in their field.

They did find that the number of schools in the survey that had implemented standards increased from the survey 10 years before by a significant amount: 59% of elementary schools compared to 19% previously that used the national standards or state standards. At the secondary level the increase was from 30% to 83%. Pufahl and Rhodes noted that a large part of this increase was due to public schools that used the standards.

Pufahl and Rhodes (2011) stated that only the respondents from the public schools were asked if there were effects from the No Child Left Behind Act on foreign language learning at their school and one third responded yes. Pufahl and Rhodes reported that comments included that there was lack of funding for foreign language because it was not a tested subject and that focus was redirected to subjects that were tested.

Pufahl and Rhodes (2011) argued that there was a discordance between what the country purports to be the goal of world citizens who speak at least one additional language besides English and the actual practice of offering the option of effective foreign language learning to students in the schools. Pufahl and Rhodes recommended more programs and funding in the schools that are more intensive and aim at proficiency in the language, providing the opportunity for students to begin earlier and have continuous instruction throughout their school career, better training for teachers, and more research on effective methods and how much language education is needed for students to become proficient.

These studies suggest that although foreign language was one of the subject areas for which the national standards were funded and developed that there still was not a full appreciation of the benefits to students of foreign language learning as it was not available to or required for all students regardless of their post-secondary study intentions. The Partnership for 21st Century Skills (P21, 2011), along with researchers and other organizations, developed a

map with guidelines for foreign language learning as well as for all the core subjects. The map outlined the cognitive benefits and skills, including social skills, which students can develop during their foreign language learning and study such as articulating ideas; collaborating; innovating; and showing initiative along with awareness and critical analysis of media, technology, and society (P21, 2011). These are skills that are seen as vitally important in our changing world (P21, 2011; Committee for Economic Development, 2006).

P21 (2007) also published on student assessment as a part of developing 21st century skills. Assessments are part of our educational system, particularly coming to the forefront with the No Child Left Behind Act of 2001 which led to more standardized testing that tests discrete point items. P21 argued that the type of testing done in United States education does not evaluate whether students can analyze and solve problems, it does not help teachers make informed decisions about how to teach in their classrooms, and it does not give a good picture of the student's school career or if their education has been effective.

P21 (2007) argued that our schools need to begin testing cognitive and social skills. Both summative and formative assessments can be used effectively. Summative assessment can be used to view if learning has occurred but with certain standards of usefulness, reliability, and fairness. Formative assessments should be used to show student achievement and reasoning, to help students shape and reflect on their learning and behaviors, and to inform future instruction for the teacher (P21, 2007).

In summary, language education has gone through key changes over the last few decades in reaction to national concern about students' global competence (Committee for Economic Development, 2006; P21, 2011). According to P21 (2007), language learning had been transformed in many cases from learning about the language to actually using the language in

different real-life applications. Research covered in this section has prompted language and non-language educators to think about performance-based assessment before deciding what and how they will teach as well as incorporating the national standards into instruction (NSFLEP, 2006). In addition, these resources promoted activities involving assessing the three modes of communication in the national standards: interpersonal, interpretive, and presentational.

Integrated Performance Assessment

The language teaching profession has encountered an array of theories and trends in practice in the past few decades. On one hand, the profession has been revitalized by the excitement generated by a renewal of interest in language learning and more creative ways to enhance learning in the classroom. On the other hand, teachers have been challenged by the need to sort through an eclectic blend of instructional ideas and to clarify their own vision of the learning/teaching process. Professionals in education rely on current research, organizations, periodicals, and other information that reaches them through staff development, membership affiliation, licensure work, and initiatives from educational entities.

One of the main goals of the ACTFL has been to support language education. This organization recommended IPA as a best practice for curriculum development to connect performance-based assessment to instruction as well as learning (Glisan et al., 2003; Adair-Hauck et al., 2006). Glisan et al. (2003) reported that the IPA concept had its origins in teachers' struggles to link the national standards and K-12 performance guidelines into their classroom instruction. In 1997, the ACTFL received federal funding to design an assessment prototype called the IPA. Glisan et al. (2003) and Adair-Hauck et al. (2006) conducted their seminal studies in IPA implementation in the classroom. This was a three-year study (1997-2000) was sponsored by the U.S. Department of Education International Research and Studies Program.

The researchers followed six pilot sites in a variety of geographic locations in middle schools and high schools of the United States. The pilot sites used authentic performance-based assessments created by the project developers with three modes of communication used in real-life situations: presentational, interpretive, and interpersonal. Chinese, French, German, Japanese, and Spanish teachers piloted the assessments. The study was organized by the ACTFL.

Glisan et al. (2003) focused on three ACTFL goals. The first goal was to design IPAs to develop a new way to assess students. The second goal was to focus on the effectiveness of IPAs in a typical classroom for measuring students' progress in meeting the communication goals of the national standards. The third goal of the IPA project, according to the researchers, was to incorporate the assessment into the foreign language classroom to inform and improve learning and instruction relative to the national standards.

Glisan et al. (2003) provided detail on the first goal area of the national standards, which is the Communication goal area. This goal consisted of three standards for communication. The IPA was used to assess students on three modes of communication. According to Glisan et al. the Interpersonal mode was intended to fulfill Standard 1.1; Standard 1.1 stated "Students engage in conversations, provide and obtain information, express feelings and emotions, and exchange opinions" (Glisan et al., 2003, p.17). The Interpretive mode works to meet Standard 1.2 which states, "Students understand and interpret written and spoken language on a variety of topics" (Glisan et al., 2003, p.17). The Presentational mode focuses on Standard 1.3. , "Students present information, concepts and ideas to an audience of listeners or readers on a variety of topics" (Glisan et al., 2003, p.17).

Glisan et al. (2003) introduced the context and purpose for the series of authentic tasks included in each IPA. They argued that the introduction provided a framework for the

assessment and illustrated how each IPA task built strength into the other culminating with an oral or written presentational task.

Glisan et al. (2003) described that in the first phase of the IPA, the teacher assessed students' performance on an interpretive task by presenting an audio/video clip or a written text. Students then answered comprehension questions based on this text. In the second phase, the Interpersonal task of the assessment, students conversed spontaneously in the target language about the same theme covered in the first stage. In the third and last phase, students gave a presentation, either oral or written, synthesizing the learning in the other two tasks. Each of the three tasks was to be scored with a rubric, based on a design by Glisan et al (2003). Students were to be familiar with these rubrics since the beginning of the unit and throughout their use in formative assessments during class time. Students also were to receive feedback after the culmination of each IPA. Glisan et al. (2003) were interested in the potential of this study to create a new model for assessing students' proficiency in the IPA in relation to the national standards. They concluded that this process of assessment was supported in their research.

In their work, Adair-Hauck et al. (2006) considered other fundamental elements in the foreign language scenario and found that development of IPA combined important concepts in best practices and research in assessment. Their second language acquisition research considered studies on how students learn a second language and at the same time how the application of this research changed teaching practices. Adair-Hauck et al. studied if there would be a positive "washback effect" from the IPA to improve instruction

Adair-Hauck et al. (2006) described how the IPA consisted of designing authentic assessments with a learning purpose related to the real-world. The students were provided with the rubric in advance and the instructors used it as a tool for feedback and provided examples of

what was expected in the final product. The researchers noted how the use of the rubric for teachers also could serve to improve instruction.

Adair-Hauck et al. (2006) found that authentic assessments embraced transparency as the students knew what was expected from them at the beginning of the unit or task. In that instance, students were informed of what they needed to do in order to attain language standards defined in the rubric; at the same time a passing grade or above becomes more transparent. The authors proposed that with this approach the learning process developed in an atmosphere of trust; trust would be built when teachers, school administrators, and parents know not only what students are doing in class but why and how their performances are going to be graded using rubrics with specific targets.

Adair-Hauck et al. (2006) pointed out that rubrics played an important role not only in the assessment process but that rubrics acted as a tool to provide feedback during the entire instruction process. The teachers started with sharing examples of past IPAs with a goal of showing what good performance might look like even before students performed assigned tasks. This included going over criteria. Feedback continued as a student worked towards attaining skills and specific goals of communication. The authors argued that students did better when they knew how they were doing and how they could improve before the assignment due date. Adair-Hauck et al (2006). stated teachers also benefited from the feedback process and activities; they could determine the direction or redirection of instruction in individual cases or as a class in general.

Adair-Hauck et al. (2006) also conducted a qualitative study during the field trials of the prototype IPA. The researchers surveyed teachers involved in the pilot study about the effect IPA would have on their own classroom instruction and assessment design going forward. The

teachers were aware of their strengths and weaknesses in both during the study; 83% said the effect would be positive on instruction and 91% would it be positive for assessment design. The researchers proposed that this data showed the desired washback effect of IPA study on teachers' perceptions. Student success in meeting the national standards was not reported. In addition, ACTFL (Glisan et al., 2003) recommended IPA as a best practice for curriculum development, as a way to connect assessment to instruction as well as learning. This study was the basis for ACTFL guidelines for IPA (Sandrock, 2010).

In 2004-2005, Glisan, Uribe, and Adair-Hauck (2007) conducted a research project to measure post-secondary students' progress towards meeting the standards for foreign language in the 21st century. The researchers implemented the IPA in a Latin American Culture and Civilization course. Glisan et al. examined how many students in the study met the Communication standard for the three modes of interpersonal, interpretive, and presentational communication and how study in secondary and post-secondary influenced performance in these three modes.

Glisan et al. (2007) described the background of the study that took place at the US Air Force Academy. The participants in the project were 29 cadets from three sections of an advanced Spanish course in Latin American culture and civilization; which covered the history of Latin America from pre-Columbian civilizations to the present and focused on current cultural issues such as societal changes, the economy, and illegal immigration. Glisan et al. (2007)described the study as follows. The course was taught by two native-speaker instructors with 15 years of language teaching experience. There were 20 males and nine females cadets in the class, with an average age of 20.3, representing various areas of the United States. The authors detailed that this was a selected group of students, with highly motivated personal and

academic levels. Only one of the participants was a native speaker of Spanish, while the others had studied Spanish an average of 1.8 years in middle school and 3.3 years in high school. Most participants had previously studied three semesters of Spanish and were considered intermediate level.

Glisan et al. (2007) explained that the IPA was implemented as a regular element of the course where both instructors and one of the researchers followed the exact same procedures. Prior to the start of the project, students completed questionnaires on language background, language perception, and foreign language motivation and anxiety; students repeated the same questionnaires after the implementation of the project. The cadets also participated in one IPA practice with all three modes of communication to ensure familiarity with the IPA format and the rubrics at the start of the project.

The interpretive mode was the first one to be in place, according to Glisan et al. (2007). Students listened to a passage from Latin American writer Carlos Fuentes' video program that addressed illegal immigration into the United States from his perspective. Students answered the interpretive mode questionnaire that they received at the beginning of the class to have time to preview the questions. In the next session, students received feedback and scores on the interpretive rubric and spent half of class time reviewing the answers to the questionnaire.

For the interpersonal mode, Glisan et al. (2007) reported that students were randomly paired up to engage in a debate about illegal immigration into the U.S. Students were prompted to use the information they had learned in class and from the interpretive activity to defend their positions, either for or against an open door to immigration. The discussion was recorded and graded using the interpersonal rubric. Glisan et al. explained that students received their individual written and overall feedback.

According to Glisan et al. (2007), the last task was presentational. It consisted of two tasks, one where students wrote a paper describing their opinions and positions about immigration using what they learned in class and from the two previous performances. The second task was to give a 2-3 minute oral presentation in front of the class about their position on immigration. Students received their individual grades and feedback using the presentation rubric for performance assessment. The authors found that the data revealed better performance in the presentational mode than in the interpretive mode. Additionally, there was a positive correlation between students' performance and number of years of Spanish study in middle school in the interpersonal mode. This confirmed earlier research findings that Glisan et al (2007) had reviewed.

Glisan et al. (2007) found a negative correlation between performance and the number of years Spanish was studied in high school. While they found this unexpected they recognized that the small number of participants was a limitation of the study.

Glisan et al. (2007) stated that the Air Force Academy research and ACTFL research suggested that U.S. high school classrooms may still be focused on grammar and form rather than communicative abilities. Another conclusion from both their study and that of Adair-Hauck et al. (2006) was the need for professional training to assist language teachers to update professional practice to connect teaching and assessment in the three modes of communication with emphasis on feedback to benefit students and instruction.

Davin, Tryon, Donato, and Hellman (2011) felt there was a need for studies of IPA with young students and studied how the IPA functioned at the elementary level with the three modes of communication. The researchers also asked if changes should be made in usage of the IPA with students at that grade level. The study was conducted in a K-8 private school in western

Pennsylvania with 48 Spanish students in fourth and fifth grades who received 15 minutes of daily instruction. The students had studied Spanish for four years. The researchers also asked what perceptions students had of the IPA. Students were familiar with the rubrics used in IPAs because they had used them during the semester previous to the beginning of the unit in the study. (The rubrics used were for novice level ACTFL IPAs, according to Davin et al. [2011]).

Davin et al. (2011) reported that after four months of instruction related to The Caribbean Sea unit, students completed their IPA tasks in nine days. They started with an interpretive listening task with a video about the beach for which the class received feedback from the teacher after individual assessments. Next was the interpersonal task where the students were paired with another student who was considered to have the same proficiency level. The researchers required this because they assumed students might not have shown what they knew if the pair was not equally matched. The students each described a picture to the other to make a comparison between the two pictures with prompts in English from the teacher. The prompts were an adaption of the IPA for this elementary class. The class received feedback from the teacher after the individual IPAs were scored. Third, according to Davin et al. was the presentational task where the student produced a written magazine article. Feedback was again given to the entire class as a whole after assessment. The assessments were designed by the researchers in conjunction with the teacher who had been trained in giving foreign language assessment. Davin et al observed the assessments in the classroom and worked with the teacher to agree on the rating. The IPAs of the individual students were collected for the researchers to study. The researchers compared the results from the different modes and also compared results of students who were studying Spanish as a second language compared to students who had additional languages. Davin et al. also surveyed the students about their language background

and their perceptions of the IPAs and the teacher was interviewed twice, before and after the IPA in the study.

Davin et al. (2011) used descriptive statistics to analyze students' performance on the three IPAs. The students had the best scores on the interpersonal task with the lowest scores on the interpretive task. The researchers reported that the students' scores were converted from the data to scores of acceptable (or better) and not acceptable: 73% of the students had unacceptable scores in the interpretive, 43% had unacceptable scores in the interpersonal, and 60% had unacceptable scores in the presentational. The researchers considered that the data showed a consistency with scores between the interpersonal and presentational modes, and that those scores were not consistent with the scores in the interpretive mode. The researchers were not surprised at that finding because they did not expect that students' ability would be equal in the different modes of communication. Students who thought the interpretive and presentational tasks were difficult did poorly on those tasks but the researchers did not find a correlation between the scores in the interpersonal task and how difficult the students perceived that task to be. Since the interpretive task is supposed to be the basis for the work in the interpersonal task, and the presentational task is supposed to build on this, Davin et al. wanted to compare the relation of the scores in the different tasks.

In the student surveys, Davin et al. (2011) found that the students understood the focus of the IPAs was content and not grammar; the students had a realization of this regardless of their performance. The students were able to articulate the knowledge they felt they were missing but perceived the IPA as a challenge or something that made their achievements stand out more than their errors and gave them a desire to fill in the gaps in their knowledge or ability to speak.

Davin et al. (2011) stated that due to the short duration of the daily lesson most of the class time was spent on interpersonal communication so this could be a reason the students performed better in the interpersonal task. Davin et al. thought that the students may have done poorly on the interpretive task because the video shown included a lot of vocabulary students had not been exposed to previously.

Davin et al. (2011) thought that it was necessary to change the language of the rubrics for elementary level so students of that age could understand the rubrics. The data indicated to the researchers that multilingual students (who spoke Spanish, English and other language(s) at home) in particular did better in the interpersonal task than the students who were currently learning Spanish as a second language and that further research was needed in that area. These researchers made the assumption that students might not have shown what they knew if the student pair in the interpersonal task was not equally matched but this was not studied.

In conclusion, Davin et al. (2011) argued that the IPA had a washback effect for the teachers and recommended further research studying this washback. They also recommended study of the effect of using IPAs over time as students continued to study the language and teachers planned instruction and assessed students assigning national standards based proficiency rating for students. The researchers recommended use of IPA at the elementary level using backward design as described by Wiggins and McTighe (2005) for curriculum design by identifying the learning outcomes in each IPA at the outset of each thematic unit.

In the same vein as the elementary students who connected their instruction to assessment, Zapata (2016) studied if university students were able to see advantages in their learning and connect the assessment process to the language instruction in their classes that used IPA. Zapata questioned university students at a private university in southern California that

moved to using the national standards in foreign language instruction using the three modes of communication (interpretive, interpersonal, and presentational) and adopted IPA as a practice for fall semester 2013. The instructors at the school, according to Zapata, were given training in the ACTFL Oral Proficiency Interview and in use of IPA. Zapata reported that there were 20 experienced instructors and six inexperienced graduate students who were trained and would be teaching the foreign language classes.

Zapata (2016) asked the students about their opinions and reflections on the use of IPA in the foreign language classroom and its effectiveness. Zapata reported that the university used a textbook in all classes to direct the topics but the substance of the class was material prepared specifically to use the three modes of communication, using authentic materials that would enable students to fulfill the assessments. Zapata described how students were informed of the purpose of the textbook and the IPAs, read and listened for the interpretive mode, and wrote a composition for the presentational mode. For the interpersonal mode, the students worked in pairs. Then Zapata administered a questionnaire both for fall and spring semester for a total of 1,236 students who answered the questionnaire.

Zapata (2016) questioned students at the novice through mid-intermediate level. The questions used from the survey followed a Likert scale method (with a range of like and dislike) focused on the format of the IPAs, if the IPAs connected to the instruction in the class, and if class activities were relevant to success in the IPAs. Zapata explained that there was also space in the questionnaire for students to elaborate or add extra comments to their answers. Data was analyzed by Zapata who suggested that overall students liked the use of IPAs but level three students were the most likely to have negative comments. Those comments, reported Zapata, stated concerns about if the student would receive a lower grade because they were more used to

traditional grammar test for which those students thought it was easier to memorize the answers and included comments that students were only trying to fulfill a requirement and not learn a language. In contrast, the researchers noted, students in level one and two most often commented on the difficulties they had with understanding when they were listening for the interpretive mode. Students on all levels saw the connection between classroom instruction and the IPAs. Zapata found that students indicated that certain instructors did not implement new curriculum in their classroom and with those instructors the students did not see the connection between learning and the IPAs.

Zapata (2016) argued that the study indicated that IPA was a valid part of foreign language instruction. Since novice learners said they had the most difficulty with the interpretive mode, the researcher said it would be to the learners' advantage, especially for the lower levels, if administrations allowed more time to be spent on preparing curricula with authentic materials. Zapata indicated that overall the students were positive in their comments about IPA. The biggest indicator of if students had a positive outlook and understood the connection with learning and the IPA, according to Zapata, was if the instructor followed through on explaining the instructional purpose of the change in curriculum and the use of IPA, and if the instructor implemented those changes in the classroom. As it was not a part of the study, the researcher did not know if lack of implementation was from teachers' lack of knowledge and training or if it was from teacher resistance to the new type of curriculum. Zapata argued that one of the biggest factors in the success of the IPA (and other educational innovations) was if and how the teacher implemented the IPA.

This section has reviewed the performance assessments using authentic tasks framed in a thematic unit linked to standards-based instruction and learning as proposed in foreign language

learning in the IPA. Adair-Hauck et al. (2006) presented data that showed the desired washback effect of IPA study on teachers' perceptions. Davin et al. (2011) showed that IPA proved to be an effective assessment of foreign language proficiency in the three modes with elementary students and Zapata (2016) with post-secondary students. Davin et al and Zapata also demonstrated that students connected classroom learning to the IPA. So although the research reviewed in this section demonstrated some effectiveness in IPA implementation and there seemed to be positive washback, Sandrock (personal communication, August 11, 2016) stated that IPA is still not fully practiced in foreign language classes even though school districts have it as part of their programs. Since school districts have adopted IPA as part of the curriculum and some of these studies touched on the effect of teachers on implementation the next section will review literature more specifically on teachers' perspectives.

Teachers' Perspectives

In the last 30 years foreign language education has incorporated key initiatives to clarify expectations in the field and to promote language learning. This section of the literature review discusses how a teacher's perceptions may affect the use of these initiatives in the classroom.

A challenge in the foreign language field related to ongoing perceptions of language instruction is the lack of experts in second language acquisition. VanPatten (2015) conducted a study of language departments in 27 colleges and universities to ask if the instruction for future language instructors incorporated the latest knowledge in language acquisition in the field. VanPatten included major research universities with doctoral programs in Spanish and French in the study. The researcher looked at tenured faculty at those institutions who would be instructing graduate students.

VanPatten (2015) found that the percentage of language experts in those institutions was significantly low and instructors at the upper division level normally had little language expertise. There were 6% of the faculty for Spanish and 2% of the faculty for French whose area of expertise was language acquisition, 48% of the universities in the study had no faculty in that area for Spanish and 78% had none for French. VanPatten found this to be a concern; he posited that the university students would not have the knowledge in language acquisition needed to inform their future instruction and curriculum.

VanPatten (2015) stated that a limitation to this study was that most Ivy League universities were not included in the data but the Ivy League universities had no experts in language acquisition. A wider range of universities and colleges were not included but the researcher stated those institutions often use the large research universities as a guide for their practice. The researcher did not include non-tenure and adjunct lecturers because those are not the professors who work with graduate students.

VanPatten (2015) surmised that the lack of expertise at this level perpetuates ideas about language learning and teaching that are not supported by new research and that this in turn means new training is not developed for students training to be instructors nor are new textbooks developed which incorporate new understanding of language acquisition.

Sercu and Castro (2005) stated that quite a few instructors have commented that they felt uncomfortable giving students true communicative activities due to curricular and time restrictions. Sercu and Castro noted that teachers worried that these activities would take time away from preparing students for exams practicing target morphology and new vocabulary. This worry was in contrast to what VanPatten (2002) posited that true acquisition could not be

happening in that type of language class, as language acquisition has been found to occur primarily when learners are engaged in true communicative activities.

The North Carolina Department of Public Instruction (NC DPI, 2001) outlined teachers' concern about oral assessment; teachers felt they did not have time for oral assessment, they did not know if they should focus on form rather than communication, and they had trouble grading the assessment. "Teachers have determined that mispronounced words, grammatical mistakes, wrong choice of words, and hesitations were assigned a point value which was deducted from the students' oral presentation, and, in doing so, they have devalued the content in favor of the form." (NC DPI, 2001, p. 6). Hence, the organization wrote, teachers can send a message to students about what is important in learning and speaking a language even when using oral assessment instead of end of chapter testing.

A qualitative study by East (2015) echoed some of these teacher concerns about oral assessment demanding too much of teachers' time. East examined teacher attitudes in oral assessment asking them in a national survey in New Zealand their opinion comparing the use of two different programs for assessment, an unscripted student to student conversation called "interact" compared to a teacher-directed teacher to student conversation called "converse". Since teachers are the ones who are actually implementing the assessments, or at times deciding not to implement them, the researcher wanted to survey their perceptions. East explained that the assessment interact is a three times yearly high-stakes assessment of spontaneous real life recorded conversation. This survey was done two years after implementation of the program. The survey was sent to only the lead teacher in each language in each school. Over 25% of teachers responded, which East considered a good response rate.

According to East (2015), teachers were anxious at the start of the program about this type of assessment being time consuming for students and for teachers, and also that students would lose interest if they felt it would take too much of their time. Teacher perception was that the student to student testing was more valid but that it was less practical than the teacher to student assessment. East also found that teachers commented that students did not like their inability to prepare for a high-stakes assessment since the conversation was supposed to be a real life and spontaneous situation.

East (2015) reported that teachers surveyed were familiar with both programs but did not have to use the assessment to participate in the survey. East found that teachers who used interact were more positive about the program than teachers who did not use it. All teachers were concerned about the practicality of interact but teachers who used the program were less negative than teachers who did not use interact. East stated that the limitations to this study were that it surveyed only teachers and the researcher did not know the reasons why 74% of the teachers did not respond. East argued that for continued success with interact it should be used within the classroom as part of instruction rather than only in its current use as a high-stakes assessment.

In a study related to teachers' actions and perceptions, Kissau and Adams (2016) studied how IPA was used in the classroom with communicative learning in the three modes (presentational, interpretive and interpersonal). To collect data, Kissau and Adams used quantitative and qualitative methods which showed what was emphasized in the classroom between the three modes used in formative and summative assessment and if this balance coincided with teachers beliefs of importance of study of each mode.

Kissau and Adams (2016) conducted their study in a K-12 private school in the southeastern United States in level one and two foreign language classes distributed between

four languages for students in grades 6-9. There were nine classes that participated in the study. The researchers reported that the school had decided to move to proficiency-based instruction using backward design. A curriculum was developed and teachers who were given training on the three modes developed the IPAs. The teachers had all previously used traditional grammar-based curriculum and testing and were on board with the new program and interested in viewing their progress in assessment through this study. The study was done in the first year of IPA usage in the school, according to Kissau and Adams.

Kissau and Adams (2016) reported that the teachers provided the assessments done for each of the modes of communication for each unit of study over the course of a semester. They found that teachers did not have to use the same assessment in their classroom as other teachers. The researchers also interviewed five of the six teachers who volunteered to be interviewed and seven random students. Students interviewed for their study had a good understanding of what the different modes of communication meant, how these were used in class, and what would be expected in the IPA.

Kissau and Adams (2016) analyzed the assessments that were uploaded to determine which modes of communication each teacher had assessed. An assessment that was determined to include more than one mode was counted individually for each mode. In analysis of the data, the researchers calculated a percentage spent on each different mode based on the total number of assessments, formative and summative, by each individual teacher.

Kissau and Adams (2016) reported that teachers said in the interviews that the interpretive listening and the interpersonal modes were the most important to focus on for novice learners but the data from the analysis of the IPAs showed that they were not putting this belief into practice. They found that overall 19% of assessments were interpretive listening, and at 13%

the lowest percentage of time overall was spent on the interpersonal assessment. Between 15% to 60% of assessments were in interpretive reading, varying widely between the classes. Kissau and Adams stated that teachers pointed out in the interviews that the interpretive reading type of assessment seemed most familiar and closest to traditional testing, so they felt more experienced with it. The teachers stated, according to the researchers, that they found the interpretive listening more difficult to implement because the instructional materials were not readily available. Teachers assumed, they reported, the best interpersonal assessment would be between teacher and student but that this would be time consuming and there was loss of class time assessing each student.

Kissau and Adams (2016) found that teachers felt inexperienced and unsure when assessing the interpersonal mode whereas the presentational writing was considered by the teachers as a more important skill for more advanced students than for students at the novice level like those in the study. However, the data they reported showed that in practice some teachers emphasized writing more and it varied from 8% to 55% among the level one and two classes. Kissau and Adams found again that teachers stated that they were more experienced evaluating writing and only assessed presentational speaking 3% to 22% of the time.

Kissau and Adams (2016) concluded that although the teachers had intentions to implement IPA, teachers did not implement it in all the same way. The researchers found that teachers felt they needed more training and that more collaboration and sharing of materials with other teachers would also save time and encourage consistency. The researchers noted that teachers said they needed more self-analysis and reflection about why they may not be implementing the IPAs as they intended. They also wanted guidance on scoring the interpersonal mode and on how to put into practice using student to student interaction for this assessment.

Kissau and Adams (2016) concluded that although the teachers were committed to the practice of IPAs that the implementation in the classroom tended to favor the modes closest to traditional practices that they had experience with and not the interpersonal mode that the teachers felt was the most important. Kissau and Adams surmised that to encourage balance in the modes there could be more technology and instructional materials made available, usage of college students training in education, and blogs for teachers to share information and materials in addition to continued training, administrative support, collaboration, data analysis, and reflection to help teachers work through long standing beliefs and practices.

Knowledge, training, time, and reflection for teachers are important pieces of implementation of new ideas and practices, according to the studies in this section. These studies have indicated that teachers have a demonstrable effect on whether policies and practices are implemented and teachers' beliefs can affect implementation (Van Patten, 2015; East, 2015). The way a teacher learned or what the teacher is accustomed to, according to East (2015) and Kissau and Adams (2016), can affect what that teacher uses in the classroom even in contrast to their stated beliefs. In this sense, Kissau and Adams felt reflection could be useful. The study by Kissau and Adams found what others had as well; teachers found themselves reverting to or putting more focus on practices they were most familiar with even though they felt they were committed to implementing IPA.

Conclusion of Literature Review

To ensure that foreign language students succeed it is important that the student population is well served; assuming language learning takes into consideration the most relevant research in language acquisition, core practices, and standards. Current research, government agencies, parents, educators, and members of the community have recognized the benefits of

language education as a critical component to our students' success in the global world (Committee for Economic Development, 2006; P21, 2011; Pufahl & Rhodes, 2011). This recognition is far from the reality as Pufahl and Rhodes (2011) reported that although the number of public high school which offered a foreign language remained steady from 1987 to 2008 the number of elementary and middle schools which offered foreign language learning dropped in this time period. The NSFLEP developed national standards for foreign language learning in 1996 and reported in 2011 that an increasing number of schools based their curriculum on those standards but few had changed the type of assessment used (ACTFL, 2011). P21 (2007) argued that assessments needed to be changed to better reflect student learning and cognitive skills. The report noted that assessments should also inform future instruction for the teacher.

According to the literature reviewed here, the 1997-2000 IPA prototype study integrated an authentic multitask assessment, standards-based connection with instruction. Students were to acquire communicative skills through real world activities framed in a thematic unit. As the seminal studies reported (Glisan et al., 2003; Adair-Hauck et al., 2006) the overarching goal was meant to optimize the reliability in assessing student progress using the standards.

Studies on IPAs have found that there was not only the need to train educators on how to use valuable tools like IPAs but there was a need for more investigation into IPA effects on teaching, learning, and assessment (Adair-Hauck et al., 2006; Glisan et al., 2007; Davin et al., 2011; Zapata, 2016). The literature highlighted that it is important to have a mechanism for more effectively informing teachers of valuable research and core practices. As Van Patten (2015) stated, the advances in language education are slow to be put into practice. Educational institutions need to provide time and training for teachers who might not have sufficient PD or might not welcome changes in the field (Abbott & Phillips, 2011). Teacher beliefs in more

traditional practices interfered with implantation of new practices (NC DPI, 2001; East, 2015). Reflection, more training, more time for collaboration and sharing of materials and assessments, and support from administration seemed to be mentioned repeatedly as important suggestions to help teachers follow through on actual implementation of practices in the classroom (East, 2015; Kissau & Adams, 2016; Sandrock, 2010).

CHAPTER 3

RESEARCH DESIGN

Several states and local districts have aligned their curriculum with the World-Readiness Standards for Foreign Language Learning in the 21st Century (ACTFL, 1999) (Appendix A) in order to build a program towards teaching for real world competence in the goal areas of Communication, Culture, Connections, Comparisons, and Communities. Research has shown the value of IPA in assessment and instruction to measure students' progress in meeting the standards for learning a foreign language and the use of the ACTFL Performance Guidelines for K-12 Learners (ACTFL, 1998) (Appendix B) to help teachers and students understand the performance outcomes for learners at various points along their language acquisition (Glisan et al., 2003).

These guidelines are represented in the Wisconsin standards for foreign language including emphasizing communication using theme-based curriculum with IPA (Keen et al., 2001). There is a local need to clarify the purpose of traditional end of chapter tests which continue to be used and ask to what extent they serve to promote Wisconsin goals and national standards.

This study examined the impact of IPA in the foreign language classroom at the high school level analyzing students' success achieving Communication Standard 1.1. This standard requires students to engage in conversations, to provide and obtain information, to express feelings and emotions, and to exchange opinions. It is compared to traditional or end of chapter grammar and vocabulary testing. Specifically, the study asks how the use of IPA compared to an end of chapter test and how this impacted high school students' success in foreign language learning as reflected in students' grades.

Data was gathered from the teacher's grade book of five classes encompassing a total of 147 secondary students in grades 9 through 12. The class was instructed using backward design for a shopping unit for six weeks. At the beginning of the unit the instructor identified for students the expectations for the end of the unit. These expectations were based on the minimal standards for foreign language learning from the ACTFL Performance Guidelines for K-12 learners, IPA rubrics, and Can-Do statements.

Setting and Participants

The study was conducted at a high school in northern Wisconsin with five classes of second year Spanish with a total of 147 students in grades 9 through 12 taught and evaluated by the same instructor during the 2015-2016 school year. The participants were comprised of 66 male and 81 female students with a variety of academic ability levels, but within 70% to 100% grade overall in Languages Arts and Math. The participants were a convenience sample that provided data accessible to the researcher. The school's enrollment was mainly comprised of Caucasian students (92%) with 8% from minorities (Black 2%, Hispanic 1%, Asian 2%, and Native American Indian 3%) according to the Public School Review website at http://www.publicschoolreview.com with reference to the 2014 National Center for Education Statistics, WI Dept. of Education source.

Instrumentation

The researcher used a quantitative approach to examine what, if any, impact the IPA had on students' success in regards to their grades and developing communication skills compared to the results obtained in the end of the chapter test for the shopping unit. Data was gathered from

the teacher's grade book for both tests within the same unit for 147 students enrolled in second year of Spanish classes.

An IPA for the interpersonal mode from the textbook Realidades (Boyle, Met, Sayers, & Wargin, 2014) from the shopping unit was used to measure students' communication skills. This assessment was worth 24 points and it was graded using a Level 2 Speaking Analytic Rubric from the Fairfax County Public Schools (Fairfax County, 2013). The rubric (Appendix C) also includes conversion charts for turning the score on performance tasks into a corresponding grade. According to the Performance Assessments for Language Students for the Fairfax County Public Schools, meeting the expectations is a "B", exceeding expectations is an "A", almost meeting expectations is a "C", and not meeting expectations is a "D" or "F". Rubrics in Fairfax County Public Schools have been in place since 1995 and have been field tested and reviewed several times to give students, teachers, and parents a measurable level of performance.

Students from the classrooms in this study were informed that the following task would need to be demonstrated by them at the end of the shopping unit: "Think about a gift you've bought. Tell your partner what you bought, for whom you bought it, and how much you paid. Then ask your partner whether he or she thinks the gift was expensive or inexpensive. Your partner will then share the same information and ask the same questions about a gift she or he bought" (Boyle et al., 2014, p. 371). The Level 2 Speaking Analytic Rubric from the Fairfax County Public Schools was used to grade the interpersonal assessment for the present study. The rubric takes into consideration six categories to grade the speaking tasks: task completion, comprehensibility, fluency, pronunciation, vocabulary, and language control. Each category has four scores ranging from 1 through 4; 1 is the lowest grade a student could obtain and is associated with not meeting the expectations, 2 with almost meeting the expectations, 3 with

meeting the expectations; and 4 is the highest grade and it is related to exceeding expectations. The rubric also provides two other charts; one chart summarizes the raw score and the other converts the raw score into a percentage, a letter grade and also gives a holistic rating. The ratings are the following: does not meet the expectations, almost meets the expectations, meets the expectations and exceeds the expectations. After converting the raw score into a percentage teachers and/or students have the option to convert percentage scores to a final grade multiplying converted percentage score with max score and then dividing the results by 100 - in case the school uses points in grade book. In fact, the school where the study was conducted uses points for grading, so the last step to convert percentage score to final points grade was necessary (Appendix C).

The end of chapter test (Appendix D) from the same textbook and in the same theme included three sections in the grammar and vocabulary part: vocabulary for shopping, direct object pronoun usage, and verb conjugation in the past tense. In the first section students had to look at the drawings and cues provided and write a sentence to say whether or not the people indicated bought the items shown in the drawing. In the second section students had to rewrite the sentences from section one but they had to replace the name of the item with the correct direct object pronoun, and in the third section students had to write what people did yesterday by looking at five pictures and using them as cues. The total score for this test was 25 points. The first section was worth 10 points: two points per sentence, one point for the item (vocabulary) and one point for the verb conjugated according to the subject (grammar). The second section was worth five points: one point for each object that was replaced by the correct direct object pronoun; and the third section was worth ten points: two points for each of five sentences, one

point for the correct conjugation of the verb and one point for the complement to the sentence to express an action in the past sentence with the corresponding vocabulary word.

Procedure

The 147 participants in this study were 9th through 12th grade students enrolled in second year Spanish classes at a high school in northern Wisconsin. They were taught and evaluated by the same foreign language instructor during the 2015-2016 school year. In the 2014-2015 academic year this school chose a new method for testing foreign language students using primarily IPAs, although some teachers kept the end of chapter test as part of the summative scores for the final grade for their students. At the start of a new theme unit the teacher experimented giving instructions and learning activities for both an IPA and end of chapter test; that unit was developed over five weeks.

The lesson planning included different communicative activities related to the shopping theme; some activities were similar to the interpersonal task that was announced at the time the new theme was introduced to the class. Students were already familiar with the rubric for that specific assessment since they have used the same rubric for previous assessments. Every week students engaged in simple conversations to talk about readings, book exercises, videos, or dialogues related to buying a gift or a necessity for them or for someone else. The teacher interacted with several students and/or groups to provide feedback right after a key activity that targeted communicative skills. Students also gave each other feedback using Level 2 analytic rubric.

The IPA utilized a recorded verbal exchange between two students who were paired by the instructor the day of the IPA. The students had to talk about an item they had purchased, where they had purchased it, and if they thought they had purchased it for a reasonable price, and

to ask and answer questions between themselves regarding the item. These were evaluated according to the ACTFL PGs for the novice-low and novice-mid levels.

Students recorded their IPA conversations and submitted their videos through Google classroom or email. Their dialogues were graded using Level 2 Speaking Analytic Rubric from the Fairfax County Public Schools. The rubric also includes conversion charts for turning the score on performance tasks into a corresponding grade (Appendix C).

The chapter test from the textbook included three sections: vocabulary for shopping, direct object pronoun usage, and verb conjugation. The total score for this test was 25 points (Appendix D). To prepare students for this test, lessons were given each week utilizing "gramactiva videos", vocabulary practice and worksheets, and book exercises to practice conjugating verbs, direct object pronouns usage, and placement, and vocabulary related to items that could be purchased as a gift for someone. Prices and students' opinion about how much a particular item cost were discussed and recorded.

Scores from both tests were converted as necessary and each were entered into the teacher's grade book for all five classes. These steps allowed the test scores to serve as the data for the comparative study.

Data Collection and Analysis

Data was collected from one teacher's grade book of five classes. The data being compared is through the scores and grades from two types of evaluations: an end of chapter test from the online textbook Realidades and an interpersonal performance assessment (IPA) from the same shopping unit. The ACTFL Proficiency Guidelines for the novice and novice-mid levels (Appendix B) were used to make sure speakers could handle the task at these levels, and to understand the potential impact on academic success and attainment of standards for

communication. The data was analyzed using a *t*-test comparing the grades from the two methods with analysis of standard deviation from the means. Also the percentage of students with a failing or unacceptable grade of F or D and the percentage with an acceptable grade of C or above on each test were examined.

CHAPTER 4

RESULTS

The study proposed to examine FL learning, comparing students' success and achievement, in particular, as it related to Interpersonal Communication Standard 1.1 which required students to engage in conversations, to provide and obtain information, to express feelings and emotions, and to exchange opinions. It compared IPA grades to traditional end of chapter testing grades as a means of measuring this Standard.

The results discussed are the from the *t*-test comparing standard deviation from the means using the two evaluation methods (Appendix E) and from the percentages of students receiving grades above and below a C grade.

Analysis of Data

The results of the testing data are significant when using a *t*-test with analysis of standard deviation. The mean from traditional end of chapter testing of the 147 students was 19.418 (SD = 5.108) and the mean from the IPA of the 147 students was 20.795 (SD = 1.224), as shown in the table below, are similar. While the means for the two methods are similar, the deviation around the mean is highly variable with the traditional end of chapter method but is not variable around the mean of the IPA. The broad variance around the one mean score indicates that the students' learning was highly variable and inconsistent in showing their learning/understanding as measured by the traditional method. The lack of variance around the IPA approach seems to indicate that the students were much more consistent in showing their learning/understanding as measured by the IPA method. The difference between the two variances is significant at a probability greater than .001.

Group Statistics

	Test	N	Mean	Std. Deviation	Std. Error Mean
Grade	Standard	147	19.4184	5.10876	.42136
	Interpersonal	147	20.7959	1.22418	.10097

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Grade	Equal variances assumed	151.815	.000	-3.179	292
	Equal variances not assumed			-3.179	162.711

Table 1. Results of t-test

The actual test grade data results show the number of students at a failing or below acceptable score in the end of chapter test (Table 2). Seventeen percent of students had a grade of F in the chapter test and 12.25% had a grade of D which means 29.25% of the students do not show success in studying Spanish for that summative test in that particular unit. The IPA scores (Table 3) show all students in the study having an acceptable level of success in Spanish communication which for the terms of this study is an A, B, or C grade: 24.5% had an A, 68% had a B, and 7.5% obtained a C. Tables 2 and 3 show the number of students in each grade range for each of the two evaluation methods. When comparing both results (Figure 1), 45% of students had higher grade in the IPA, their grades went up 1 - 3 grades; 31% had same grade; and 24% had lower grade, but not a failing grade. Of this last group, 92% of those went down one grade (86% from A to B, 8% went down 2 grades from A to C, 6% down from B to C).

Table 2

CHAPTER TEST	
Letter Grade (Point Range out of 25)	# of Students in Range (out of 147 total)
A (90% to 100%)	56
B (80% to 89%)	35
C* (70% to 79%)	13
D (60% to 69%)	18
F (59% and under) *Passing Grade is a “C” or above	25

Table 2. Scores for Chapter Test

Table 3

INTEGRATED PERFORMANCE ASSESSMENT	
Letter Grade (Point Range out of 24)	# of Students in Range (out of 147 total)
A (90% to 100%)	36
B (80% to 89%)	100
C* (70% to 79%)	11
D (60% to 69%)	0
F (59% and under) *Passing Grade is a “C” or above	0

Table 3. Scores for Integrated Performance Assessment

Figure 1

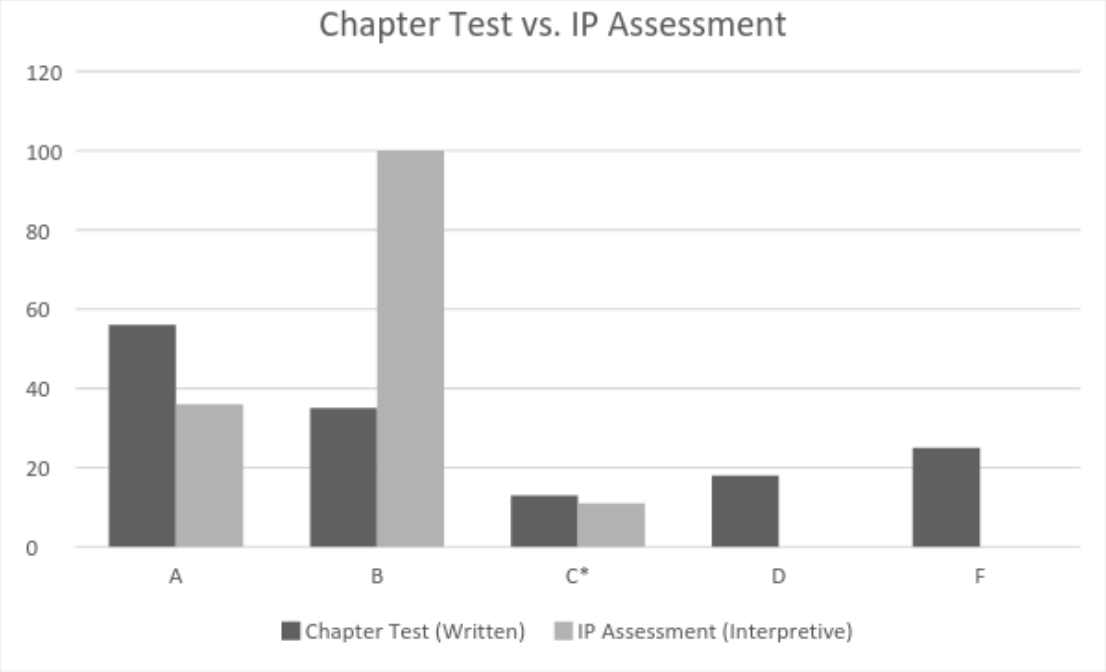

Figure 1. Comparison between Chapter test and Integrated Performance Assessment Percentage/Letter Grade

Conclusion of Results

The difference between the two standard deviations was significant and seemed to indicate that the IPA scores show significantly more consistency in evaluating students' knowledge as described in Standard 1.1.The traditional end of chapter scores seemed to show broader inconsistency in evaluating students' knowledge using the more traditional scoring method. The students' grades appeared to show that students had more success in their learning with the IPA. In contrast, the data from the end of chapter test which tests grammar and vocabulary in the unit did not demonstrate as well the knowledge that the students had gained or their ability to communicate in Spanish, as was shown in the IPA.

CHAPTER 5

DISCUSSION

The ACTFL (2012c) recommends IPA as a best practice for curriculum development as a way to connect assessment to instruction and learning. The IPA was developed as a method for teachers to link the national standards and K-12 performance guidelines into their instruction. Although it has been embraced by states and school district administrations around the country there has been difficulty with wide implementation in the classroom, as Sandrock, (ACTFL Department of Education Director; personal communication, August 9, 2016), confirmed.

The main purpose of the study was to compare the use IPA with standard end of the chapter tests to assess high school students' foreign language communication learning in a school district in northern Wisconsin. It compared two approaches for assessing foreign language learning through comparison of the students' grades from IPA to the end of chapter test grades. In the second year of implementation of the IPA in the school district there was a need to clarify if the continued use of one or both of these assessment measures served to promote the district's goals and standards and allowed the students to achieve the grade level that represents their skills in the language.

Findings and Interpretations

The students' grades showed less distribution or range of scores which implies better conversational use of the language which was the desired student outcome. The difference in the standard deviation was significant and indicated that the IPA scores showed consistency in evaluating students' knowledge and the traditional end of chapter testing was highly inconsistent in evaluating students' knowledge. One can draw from the data that the end of chapter test, which tests grammar on past tense, direct object pronouns, and vocabulary (items to buy in a

store) in the unit, does not demonstrate as fully the knowledge that the students had gained or their ability to communicate in Spanish that was shown in the IPA.

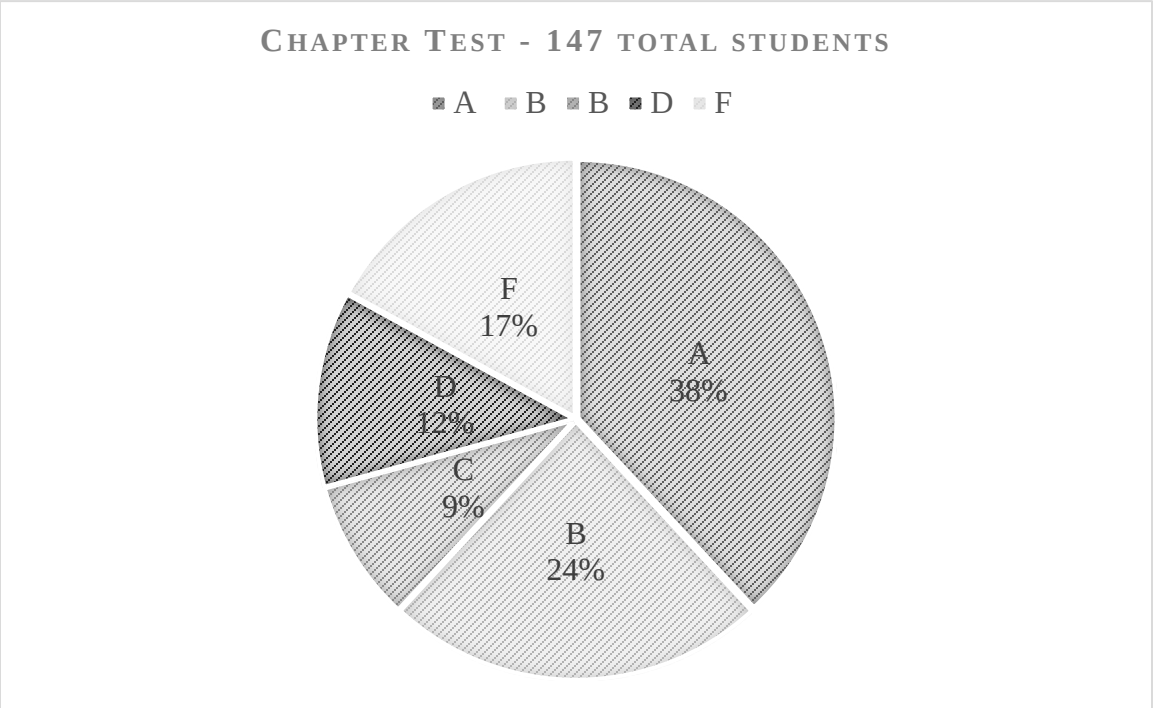

Figure 2. Letter Grade Percentage in Performance Assessment

Expectations for the IPA were based on the minimal standards for foreign language learning from the ACTFL Performance Guidelines for K-12 learners, IPA rubrics, and Can-Do statements. The researcher hypothesized that the students would have better outcomes and more accurate evaluation in accordance with national and Wisconsin standards in the IPA results compared to traditional end of chapter testing results. This hypothesis is borne out.

The IPA scores showed all students in the study having an acceptable level of success in Spanish communication, which for the terms of this study is an A, B, or C grade. The quantitative results indicated that the end of chapter exam was not an accurate reflection of the standards or of the knowledge that students had acquired and were capable of using to

communicate in Spanish. Students who failed the test based on the end of chapter exam actually had a grasp of the communication skills required at their level, as shown in the IPA and in the ACTFL Proficiency Guidelines which guide students' knowledge and skills toward language acquisition at their particular level. The IPA was more like a "real" conversation that students had about shopping for a gift for someone. This can transfer to a situation that they could have outside the classroom. Students demonstrated that they could use the grammar and vocabulary of the lesson in their dialogues at the ACTFL novice-low to novice-mid levels.

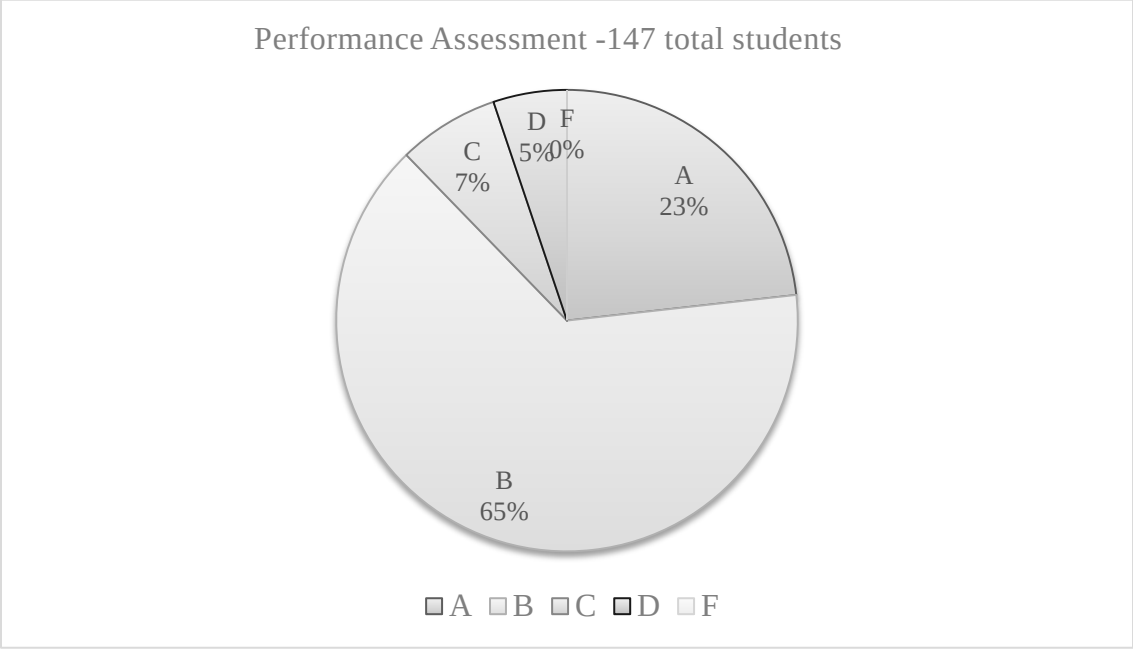

Figure 3: IPA Performance Assessments

It should be noted that the final grade in a course affected the ability to move on to upper level courses. Continued foreign language study requires at least a C grade. The low and failing grades can affect student motivation and ability to continue with study in the foreign language in the high school.

The class was instructed using backward design. Adair-Hauck et al. (2006), Sandrock (2010) and Zapata (2016) conducted their IPA studies using backward design as well. When incorporating the backward design of performance tasks into the IPA process, students know up front what is expected from them, daily activities may make sense because students are aware of the learning targets and their focus is on communication, a more collaborative learning environment may be created, and students may contribute to their own learning (Sandrock, 2010). Davin et al. (2011) reported elementary students understood the focus of the IPAs was content rather than grammar, and the students could articulate the knowledge they needed to communicate and perceived that the IPA made their achievements stand out more than their errors. It gave students a desire to fill in the gaps in their knowledge. Zapata (2016) reported university students on all levels saw the connection between classroom instruction and the IPAs.

Schools which continue to use chapter tests as the only method of evaluation could be failing students that are learning what the standards recognize as goals in foreign language learning because of the focus on chapter tests. This can send a wrong message that communication is not attainable. The NC DPI (2001) argued that students received the wrong message about their ability to learn the language when grammar and vocabulary were the focus. This could be applicable to the continued use of chapter tests in the study school. Glisan et al. (2007) reported that 62% of post-secondary students explained that their foreign language courses at the high school level focused on memorization of grammatical structures and vocabulary: 38% reported that there was a balance among listening, reading and writing, and cultural understanding: and only 3% of students reported that communication was the main focus of study.

Another conclusion from Glisan et al. (2007), Davin et al. (2011), Kissau and Adams (2016) that connected with the present project, was the continued need for PD for language teachers to connect teaching and assessment in the three modes of communication with emphasis on feedback to benefit students and instruction. East (2015) found teachers who did not attempt to use student to student interpersonal evaluation had a more negative view of it than those teachers who did. Zapata (2016) found students in classes with teachers who did not use the new curriculum had a more negative view of the IPA than students in classes where the teachers not only embraced the new curriculum but incorporated IPA in instruction. Studies found that it was had to let go of what was previous practice, whether it was university professors preparing future foreign language teachers (VanPatten, 2015) or elementary level language teachers (Zapata, 2016). Teachers repeatedly affirmed the need for support to change practices – whether through professional development, time for collaboration and sharing of materials, or guidance and practice with those with expertise.

Limitations

Limitations to this study are that it was of short duration so there was limited data. The population for this study was limited as this study was done with one five-week unit with one teacher in a single school district in northern Wisconsin, the student population was not ethnically diverse, and students enrolled in this class needed to have minimum grade average of C or better in their Language Arts and Math grades in order to take or continue taking foreign language as an elective class. Another limitation was that the teacher did not have extensive experience in implementation of IPA since this was only the second year it had been used in the district.

Recommendations for Further Research

The language teaching profession has been challenged by the many theories and trends in practice in the past few decades. On one hand, the profession has been revitalized by the excitement generated by a renewal of interest in language learning and more creative ways to enhance learning in the classroom. On the other hand, teachers have been challenged by the need to sort through an eclectic blend of instructional and ideas in the field and to clarity for their own vision of the learning/teaching process. The lack of funding and emphasis by states on other areas has further undermined progress. Further research is needed to determine to what extent implementation of IPA has been successful and what are the best practices. It has now been more than 10 years since the 3-year IPA seminal studies. Further research could include doing a larger but similar study or studying the usage of IPA without relation to traditional testing with teachers who have more experience in IPA or which includes ongoing training before the study is conducted. Further research could also be done on the most effective student to student pairing for the interpersonal IPA.

Conclusion

The research attempted to examine if students could not only show an increase in their learning with IPA but that they would have more opportunity and motivation to continue in upper level study. People often define success in foreign language learning by their oral communication skills. A student needs motivation to keep learning and according to Sandrock (2010) students' motivation and success will increase when students understand what is expected of them in advance in the planned assessment. Methods for instruction and performance assessment suggested by the ACTFL aim to promote student motivation and learning (Sandrock, 2010).

Over the last few years the knowledge of the foreign language teaching and learning process has gone through fundamental changes (VanPatten, 2016). Professionals in education may lack ongoing professional training and may not have a clear understanding of the implications and reasoning behind standards driven instruction (Abbott & Phillips, 2011). The national standards and 21st century skills are a major component in planning instruction (Sandrock, 2010). Teacher reflection seems to be an important element in the process of implementation of new practices in the classroom (Kissau & Adams, 2016).

This research study was important since the district adopted IPA, making new assessment practices a critical issue within the department. This process helped identify answers to the research question that examined the learning and assessment processes that will affect hundreds of students directly. While it is important to put into practice those techniques shown to be most effective, it is also important to provide PD so that instructors can be comfortable with the change. These findings can serve as a resource for colleagues who might have experienced the same problem or who might have difficulty moving away from the traditional testing which does not seem to reflect standards or the path to language acquisition. Many teachers find themselves in this situation (Abbott & Phillips, 2011).

IPA is not the only answer, but is one to implement or continue implementing to help to target the number of students who dropout and have failing grades, and the number of students who do not continue studying a language beyond two years in the program. Carefully implemented IPA may allow students to become active participants in their own learning process, acquiring the language, and practicing critical thinking skills simultaneously. The research and review of the literature has shown this is an efficient way to motivate students to take charge of their learning.

REFERENCES

Abbott, A. & Phillips, J. (2011). *A decade of foreign language standards: Influence, impact, and future directions*. Alexandria, VA: ACTFL. Retrieved from: https://www.actfl.org/sites/default/files/national-standards-2011.pdf

Adair-Hauck, B., Glisan, E., Koda, K., Swender, E., & Sandrock, P. (2006). The integrated performance assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals, 39*, 359–382.

Adair-Hauck, B., Glisan, E., & Troyan, F. (2013). *Implementing integrated performance assessment*. Alexandria, VA: ACTFL.

American Council on the Teaching of Foreign Languages (1998). *ACTFL Performance guidelines for k-12 learners.* Yonkers, NY: Author.

American Council on the Teaching of Foreign Languages (1999). *ACTFL Standards for foreign language learning in the 21st century.* Yonkers, NY: Author.

American Council on the Teaching of Foreign Languages (2011). *A decade of foreign language standards: Influence, impact, and future directions: Survey results*. Alexandria, VA: Author. Retrieved from: https://www.actfl.org/sites/default/files/pdfs/StandardsImpactSurvey_FINAL.pdf

American Council on the Teaching of Foreign Languages (2012a). *ACTFL proficiency guidelines.* Retrieved from: https://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf

American Council on the Teaching of Foreign Languages (2012b). *NCSSFL-ACTFL Can-Do statements: Progress indicators for language learners*. Alexandria, VA: Author.

Retrieved from: http://www.actfl.org/publications/guidelines-and-manuals/ncssfl-actfl-can-do-statements

American Council on the Teaching of Foreign Languages (2012c). *ACTFL performance descriptors for language learners*. Alexandria, VA: Author.

American Council on the Teaching of Foreign Languages (2015). *World-readiness standards for learning languages*. Alexandria, VA: Author.

Boyle, P. P., Met M., Sayers, R. S., Wargin, C. E. (2014). *Realidades*. Boston: Pearson Prentice Hall.

Clementi, D. & Terrill, L. (2013). *The keys to planning for learning: Effective curriculum, unit, and lesson design*. Alexandria, VA: ACTFL.

Committee for Economic Development (2006). *Education for global leadership: The importance of international studies and foreign language education for U.S. economic and national security*. Washington, D.C.: Author.

Committee for Economic Development (n.d.). *Foreign language enrollments in K-12 public schools: Are students prepared for a global society?* Retrieved from: https://www.ced.org/pdf/actfl-k12-foreign-language-for-global-society.pdf

Davin, K., Troyan, F., Donato, R., & Hellman, A. (2011). Research on the integrated performance assessment in an early foreign language learning program. *Foreign Language Annals*, *44*(1), 605-625.

East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, *32*(1), 101-120.

Fairfax County. (2013). *Level 2 speaking tasks analytic rubric*. Retrieved from: http://www2.fcps.edu/is/worldlanguages/pals/documents/Level2AnalyticSpeaking.pdf

Glisan, E. W., Adair-Hauck, B., Koda, K., Sandrock, P., & Swender, E. (2003). *ACTFL integrated performance assessment*. Yonkers, NY: ACTFL.

Glisan, E. W., Uribe, D., & Adair-Hauck, B. (2007). Research on integrated performance assessment at the post-secondary level: Student performance across the modes of communication. *Canadian Modern Language Review*, *64*(1), 39–68.

Keen, J., Grady, S., & Sandrock, P. (2001). *Wisconsin's Model Academic Standards for Foreign Language* (Rev. ed.). Milwaukee, WI: Wisconsin Department of Public Instruction.

Kissau, S. & Adams, M. J. (2016). Instructional decision making and IPAs: Assessing the modes of communication. *Foreign Language Annals, 49*(1), 105-123.

Moeller, A. J. & Catalano, T. (2015). Foreign language teaching and learning. *International Encyclopedia of Social & Behavioral Sciences* (2nd ed.). Lincoln, NE: Faculty Publications: Department of Learning and Teacher Education.

Moeller, A. J., Cresswell, J. W., & Saville, N. (Eds.) (2016). *Second language assessment and mixed methods research*. Cambridge: Cambridge University Press.

National Center for Education Statistics: School Directory Information. Retrieved from: http://nces.ed.gov/ccd/schoolsearch

National Standards in Foreign Language Education Project (1996). *Standards for foreign language learning: Preparing for the 21st century*. Lawrence, KS: Allen Press

National Standards in Foreign Language Education Project (2006). *Standards for foreign language learning in the 21st century* (3rd ed.). Lawrence, KS: Allen Press.

North Carolina Department of Public Instruction (2001). *Oral language assessment in the foreign language class (planning, conducting, managing): The possible dream).* Raleigh, NC: Author.

Partnership for 21st Century Skills (2007). *21st century skills assessment: A partnership for 21st century skills e-paper*. Retrieved from: http://www.p21.org/storage/documents/21st_Century_Skills_Assessment_e-paper.pdf

Partnership for 21st Century Skills (2011). *21st century skills map*. Washington, DC: Author. Retrieved from: https://www.actfl.org/sites/default/files/pdfs/21stCenturySkillsMap/p21_worldlanguagesmap.pdf

Pufahl, I. & Rhodes, N. (2011). Foreign language instruction in U.S. schools: Results of a national survey of elementary and secondary schools. *Foreign Language Annals*, *44*(2), 258-288.

Sandrock, P. (2010). *The keys to assessing language performance*. Alexandria, VA: ACTFL

Sercu, L. & Castro, P. (2005). *Foreign language teachers and intercultural competence*. Clevedon, England: Multilingual Matters.

VanPatten, B. (2002). *From input to output*. San Francisco: McGraw-Hill.

VanPatten, B. (2015). Who are the Experts? *Hispania*, *98*(1), 2-13. Johns Hopkins University Press: doi: 10.1353/hpn.2015.0011

Vélez-Rendón, G. (2002). Second language teacher education: A review of the literature. *Foreign Language Annals. 35*(4), 457-467.

Wiggins, G., & McTighe, J. (2005) *Understanding by design* (2nd ed.). Alexandria, VA: Association and Supervision of Curriculum Development. Retrieved from: http://www.edutopia.org/pdfs/resources/wiggins-mctighe-backward-design-why-backward-is-best.pdf

Zapata, G. C. (2016). University students' perceptions of integrated performance assessment and the connection between classroom learning and assessment. *Foreign Language Annals, 48*(1), 93-104. doi: 10.1111/flan.12176

APPENDICES

Appendix A

WORLD-READINESS STANDARDS FOR LEARNING LANGUAGES

GOAL AREAS	STANDARDS		
COMMUNICATION Communicate effectively in more than one language in order to function in a variety of situations and for multiple purposes	**Interpersonal Communication:** Learners interact and negotiate meaning in spoken, signed, or written conversations to share information, reactions, feelings, and opinions.	**Interpretive Communication:** Learners understand, interpret, and analyze what is heard, read, or viewed on a variety of topics.	**Presentational Communication:** Learners present information, concepts, and ideas to inform, explain, persuade, and narrate on a variety of topics using appropriate media and adapting to various audiences of listeners, readers, or viewers.
CULTURES Interact with cultural competence and understanding	**Relating Cultural Practices to Perspectives:** Learners use the language to investigate, explain, and reflect on the relationship between the practices and perspectives of the cultures studied.	**Relating Cultural Products to Perspectives:** Learners use the language to investigate, explain, and reflect on the relationship between the products and perspectives of the cultures studied.	
CONNECTIONS Connect with other disciplines and acquire information and diverse perspectives in order to use the language to function in academic and career-related situations	**Making Connections:** Learners build, reinforce, and expand their knowledge of other disciplines while using the language to develop critical thinking and to solve problems creatively.	**Acquiring Information and Diverse Perspectives:** Learners access and evaluate information and diverse perspectives that are available through the language and its cultures.	
COMPARISONS Develop insight into the nature of language and culture in order to interact with cultural competence	**Language Comparisons:** Learners use the language to investigate, explain, and reflect on the nature of language through comparisons of the language studied and their own.	**Cultural Comparisons:** Learners use the language to investigate, explain, and reflect on the concept of culture through comparisons of the cultures studied and their own.	
COMMUNITIES Communicate and interact with cultural competence in order to participate in multilingual communities at home and around the world	**School and Global Communities:** Learners use the language both within and beyond the classroom to interact and collaborate in their community and the globalized world.	**Lifelong Learning:** Learners set goals and reflect on their progress in using languages for enjoyment, enrichment, and advancement.	

Appendix B

NCSSFL-ACTFL Can-Do Statements

Interpersonal Communication

NOVICE LOW

I can communicate on some very familiar topics using single words and phrases that I have practiced and memorized.

I can greet my peers.

- ☐ I can say hello and goodbye.
- ☐ I can _____

I can introduce myself to someone.

- ☐ I can tell someone my name.
- ☐ I can _____

I can answer a few simple questions.

- ☐ I can respond to yes/no questions.
- ☐ I can answer an either/or question.
- ☐ I can respond to *who*, *what*, *when*, *where* questions.
- ☐ I can _____

NOVICE MID

I can communicate on very familiar topics using a variety of words and phrases that I have practiced and memorized.

I can greet and leave people in a polite way.

- ☐ I can say hello and goodbye to someone my age or younger.
- ☐ I can say hello and goodbye to my teacher, professor, or supervisor.
- ☐ I can say hello and goodbye to an adult.
- ☐ I can say hello and goodbye to a person I do not know.
- ☐ I can _____

I can introduce myself and others.

- ☐ I can introduce myself and provide basic personal information.
- ☐ I can introduce someone else.
- ☐ I can respond to an introduction.
- ☐ I can _____

I can answer a variety of simple questions.

- ☐ I can answer questions about what I like and dislike.
- ☐ I can answer questions about what I am doing and what I did.
- ☐ I can answer questions about where I'm going or where I went.
- ☐ I can answer questions about something I have learned.
- ☐ I can _____

I can make some simple statements in a conversation.

- ☐ I can tell someone what I am doing.
- ☐ I can say where I went.
- ☐ I can say whom I am going to see.
- ☐ I can express a positive reaction, such as "Great!"
- ☐ I can _____

I can ask some simple questions.

- ☐ I can ask *who*, *what*, *when*, *where* questions.
- ☐ I can ask questions about something that I am learning.
- ☐ I can _____

I can communicate basic information about myself and people I know.

- ☐ I can say my name and ask someone's name.
- ☐ I can say or write something about the members of my family and ask about someone's family.
- ☐ I can say or write something about friends and classmates or co-workers.
- ☐ I can _____

I can communicate some basic information about my everyday life.

- ☐ I can give times, dates, and weather information.
- ☐ I can talk about what I eat, learn, and do.
- ☐ I can talk about places I know.
- ☐ I can ask and understand how much something costs.
- ☐ I can tell someone the time and location of a community event.
- ☐ I can _____

Appendix C

Level 2 Speaking Tasks
Analytic Rubric

Task Completion

1 Minimal completion of the task and/or responses frequently inappropriate.

2 Partial completion of the task; responses mostly appropriate yet undeveloped.

3 Completion of the task; responses appropriate and adequately developed.

4 Superior completion of the task; responses appropriate and with elaboration.

Comprehensibility

1 Responses barely comprehensible.

2 Responses mostly comprehensible, requiring interpretation on the part of the listener.

3 Responses comprehensible, requiring minimal interpretation on the part of the listener.

4 Responses readily comprehensible, requiring no interpretation on the part of the listener.

Fluency

1 Speech halting and uneven with long pauses or incomplete thoughts.

2 Speech choppy and/or slow with frequent pauses; few or no incomplete thoughts.

3 Some hesitation but manages to continue and complete thoughts.

4 Speech continuous with few pauses or stumbling.

Pronunciation

1 Frequently interferes with communication.

2 Occasionally interferes with communication.

3 Does not interfere with communication.

4 Enhances communication.

Vocabulary

1 Inadequate and/or inaccurate use of vocabulary.

2 Somewhat inadequate and/or inaccurate use of vocabulary and too basic for this level.

3 Adequate and accurate use of vocabulary for this level.

4 Rich use of vocabulary.

Language Control

1 Inadequate and/or inaccurate use of basic language structures.

2 Emerging use of basic language structures.

3 Emerging control of basic language structures.

4 Control of basic language structures.

Name _____

Task Completion	½	1	1½	2	2½	3	3½	4
Comprehensibility	½	1	1½	2	2½	3	3½	4
Fluency	½	1	1½	2	2½	3	3½	4
Pronunciation	½	1	1½	2	2½	3	3½	4
Vocabulary	½	1	1½	2	2½	3	3½	4
Language Control	½	1	1½	2	2½	3	3½	4

Raw Score: _____ **/ 24**

RAW SCORE	PERCENT	GRADE	RATING	RAW SCORE	PERCENT	GRADE	RATING
24	100%	A	EE	12	73.0%	C	AME
23.5	98.3%	A	EE	11.5	72.0%	C-	AME
23	96.6%	A	EE	11	71.0%	C-	AME
22.5	94.9%	A	EE	10.5	70.1%	C-	AME
22	93.4%	A	EE	10	69.0%	D+	DNME
21.5	91.9%	A-	EE	9.5	67.9%	D+	DNME
21	90.5%	A-	EE	9	66.7%	D	DNME
20.5	89.2%	B+	ME	8.5	65.4%	D	DNME
20	88.0%	B+	ME	8	64.0%	D	DNME
19.5	86.8%	B	ME	7.5	62.7%	F	DNME
19	85.7%	B	ME	7	61.2%	F	DNME
18.5	84.7%	B	ME	6.5	59.6%	F	DNME
18	83.7%	B	ME	6	57.9%	F	DNME
17.5	82.7%	B-	ME	5.5	56.1%	F	DNME
17	81.8%	B-	ME	5	54.3%	F	DNME
16.5	80.8%	B-	ME	4.5	52.2%	F	DNME
16	80.0%	B-	ME	4	50.1%	F	DNME
15.5	79.1%	C+	AME	3.5	47.9%	F	DNME
15	78.3%	C+	AME	3	45.5%	F	DNME
14.5	77.4%	C+	AME	2.5	42.9%	F	DNME
14	76.6%	C	AME	2	40.3%	F	DNME
13.5	75.7%	C	AME	1.5	37.4%	F	DNME
13	74.8%	C	AME	1	34.5%	F	DNME
12.5	73.9%	C	AME	.5	31.6%	F	DNME

EE: Exceeds Expectations; **ME**: Meets Expectations;
AME: Almost Meets Expectations; **DNME**: Does Not Meet Expectations

Converted % Score: _____ **%**

If you use points in your grade book, use Conversion Chart B or the following formula:

$$\frac{\text{converted \% score} \times \text{max score}}{100} = \text{student points}$$

(To divide by 100 move the decimal point two places to the left.)

FINAL GRADE: _____

©2013 FOREIGN LANGUAGE PROGRAM OF STUDIES, FAIRFAX COUNTY PUBLIC SCHOOLS

Appendix D

Grammar Test

Realidades 1 Nombre _____ Hora _____

Capítulo 7B Fecha _____ Examen 7B, Page 1

EXAMEN DEL CAPÍTULO, 7B

PARTE I: Vocabulario y gramática en uso

A. Look at the drawings and the cues provided. Write a sentence to say whether or not the people indicated bought the items shown in the drawings.

Modelo tú / sí

Tú compraste las botas .

1. yo / sí

2. nosotros / no

3. Uds. / no

4. ella / sí

5. ellos / no

B. Now, read the sentences you wrote for Part A, and replace the name of the item shown with the correct direct object pronoun. Rewrite the complete sentence on your answer sheet.

Modelo *Tú las compraste .*

© Pearson Education, Inc. All rights reserved.

C. Look at the items shown. Then, use the cues to say what the people indicated did yesterday.

1. Luisa / comprar

2. yo / buscar

3. nosotros / escuchar *or ... a* →

4. yo / jugar

5. Sebastián / sacar *or* →

© Pearson Education, Inc. All rights reserved.

Realidades 1 Nombre [redacted] Hora 7

Capítulo 7B Fecha 4/26/16 Hoja de respuestas 7B, Page 1

HOJA DE RESPUESTAS

*25/25

PARTE I: Vocabulario y gramática en uso

A. (10 / 10 puntos)

1. Yo compré los guantes.
2. nosotros no compramos los anteojos de sol.
3. ustedes no compraron el llavero.
4. Ella compró la carrera.
5. Ellos no compraron la corbata.

B. (5 / 5 puntos)

1. Yo los compré.
2. nosotros no los compramos.
3. ustedes no lo compraron.
4. Ella la compró.
5. Ellos no la compraron.

C. (10 / 10 puntos)

1. Ayer luisa compró unas arpas.
2. Ayer yo busqué un anillo.
3. Ayer nosotros escuchamos a la profesora.
4. Ayer yo jugué el tenis.
5. Ayer sebastián sacó la basura.

¡Hay más!

© Pearson Education, Inc. All rights reserved.

Appendix E
t-Scores Data

```
T-TEST GROUPS=Test(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Grade
  /CRITERIA=CI(.95).
```

T-Test

Notes

Output Created		17-AUG-2016 13:54:27
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	294
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST GROUPS=Test(1 2) /MISSING=ANALYSIS /VARIABLES=Grade /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.05

Group Statistics

	Test	N	Mean	Standard Deviation	Std. Error Mean
Grade	Standard	147	19.4184	5.10876	.42136
	Interpersonal	147	20.7959	1.22418	.10097

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means					

		F	Sig.	t	df					
Grade	Equal variances assumed	151.815	.000	-3.179	292					
	Equal variances not assumed			-3.179	162.711					

Independent Samples Test

		t-test for Equality of Means				
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	
Grade	Equal variances assumed	.002	-1.37755	.43329	-2.23032	
	Equal variances not assumed	.002	-1.37755	.43329	-2.23315	

Independent Samples Test

		t-test for Equality of Means
		95% Confidence Interval of the Difference
		Upper
Grade	Equal variances assumed	-.52478
	Equal variances not assumed	-.52195