DEPICTING WASHBACK IN THE INTERMEDIATE SPANISH LANGUAGE CLASSROOM: A DESCRIPTIVE STUDY OF TEACHER’S INSTRUCTIONAL BEHAVIORS AS THEY RELATE TO TESTS

DISSERTION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Lisa A. Fournier-Kowaleski, M. Ed.

The Ohio State University
2005

Dissertation Committee:

Alan Hirvela Ph.D., Adviser
Donna Reseigh Long, Ph.D.
Janice L. Macián, Ph.D.

Approved by

Adviser
College of Education
ABSTRACT

Education relies heavily on testing to make predictions about learner achievement, skill level, and future success. Tests are also a means to hold learners accountable. In recent years, the matter of accountability has become the focus of educational concerns which has only placed further emphasis on the importance of testing in America’s educational institutions. Although educators frequently claim that they do not want to teach to a test, the reality is that every educator wants his/her students to be successful. Decision makers, teachers, and students equate this success in large part with high test scores, resulting in classroom instruction that is reflective of test practices and/or expectations.

The effect that testing has on teaching and instruction is a phenomenon referred to as washback. The purpose of this qualitative study was to describe the washback behaviors of teachers, and to a lesser extent, students, in the low stakes testing environment of an intermediate level Spanish class. The study followed a group of five teachers over a period of two quarters, teaching an intermediate level Spanish course. The results detail washback behaviors exhibited by these teachers during the first quarter of teaching Spanish 104 for the first time, and then in the subsequent quarter teaching the course for the second time. Teachers’ instructional behaviors changed during the second time teaching the course as a result of increased knowledge of the test. Factors such as
past teaching experience and teacher beliefs were also found to influence changes in teachers’ instructional behaviors.

Secondary factors of interest such as student study habits, student perceptions of their teacher’s behaviors, and the effect of tests were also investigated. The students’ study and classroom behaviors appeared to change to correspond to changed behaviors exhibited by the students’ teacher. The student data serves as a potential source for further investigation into the washback phenomenon.
ACKNOWLEDGEMENTS

I wish to thank my advisor, Alan Hirvela, for intellectual support, and encouragement which made this thesis possible, and for his patience in correcting the various drafts of this document. His commitment to the Foreign and Second Language Program in Education at The Ohio State University is commendable. I am fortunate to have had him for my advisor.

I also thank Donna Reseigh Long and Janice Lynn Macián for their devoted participation in this research, and their commitment to helping me produce a document that accurately and concisely represents this endeavor. Each devoted tremendous time to the various drafts of this document, providing extensive feedback and focused guidance.

I am grateful to the Department of Spanish and Portuguese at the Ohio State University for the opportunities that have been offered to me during the past three years. This research project would not have been possible without the support and guidance of this department.
VITA

1996……………………………………………………………M.A. Educational Leadership and Management, Fitchburg State College

1990-2000………………………………………… Spanish Teacher, Curriculum Liaison Wachusett Regional High School, Holden, MA

2000-2002………………………………………… Curriculum Specialist for Foreign Languages and ESL, Wachusett Regional School District, Holden, MA

2002-2005………………………………………… Graduate Teaching Associate, Department of Spanish and Portuguese, The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Education
Minor Fields: Technology in Foreign Language Teaching; Spanish Language Pedagogy
# TABLE OF CONTENTS

Abstract .......................................................................................................................... ii

Acknowledgements ......................................................................................................... iv

Vita.................................................................................................................................. v

List of Tables .................................................................................................................... xi

List of Figures .................................................................................................................... xiii

Chapters:

1. Introduction ..................................................................................................................... 1
   1.1 Testing in Public Education ....................................................................................... 1
   1.2 Foreign Languages ..................................................................................................... 2
   1.3 Washback .................................................................................................................... 5
   1.4 Washback in Foreign Languages ............................................................................. 5
   1.5 Statement of the Problem ....................................................................................... 8
   1.6 Significance of the Study ....................................................................................... 11
   1.7 Research Questions ............................................................................................... 13
   1.8 Methodology ............................................................................................................ 14
   1.9 Definition of Key Terms ....................................................................................... 15
   1.10 Assumptions of the study .................................................................................... 17
   1.11 Limitations of the study ....................................................................................... 18

2. Literature Review ............................................................................................................ 20
   2.1 Historical Perspective ............................................................................................. 20
   2.2 Language Teaching and Test Design ....................................................................... 25
      2.2.1 Testing ‘Communicative Ability’ ........................................................................ 26
      2.2.2 Learning Theory ............................................................................................... 27
   2.3 Influences on Testing in the Communicative Foreign Language Classroom .......... 28
   2.4 Performance Testing ............................................................................................... 30
      2.4.1 Performance Test Design and Evaluation ......................................................... 31
      2.4.2 Validity and Reliability in Foreign Language Testing ....................................... 34
   2.5 The Influence of Tests on Instruction: Washback .................................................. 36
   2.6 Washback Research ............................................................................................... 38
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7</td>
<td>Washback Research in High Stakes Testing</td>
<td>41</td>
</tr>
<tr>
<td>2.8</td>
<td>Washback Research on Tests to Drive Innovation and Change</td>
<td>44</td>
</tr>
<tr>
<td>2.9</td>
<td>A Future for Washback Research in the United States</td>
<td>46</td>
</tr>
</tbody>
</table>

3. Methods

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Design and Procedures</td>
<td>51</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Research Site</td>
<td>51</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Participants</td>
<td>51</td>
</tr>
<tr>
<td>3.2.3</td>
<td>The Researcher</td>
<td>53</td>
</tr>
<tr>
<td>3.3</td>
<td>Intermediate Spanish 104</td>
<td>54</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Course Description</td>
<td>54</td>
</tr>
<tr>
<td>3.3.2</td>
<td>The Spanish 104 Testing Program</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>Instrumentation</td>
<td>58</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Observations</td>
<td>59</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Interviews</td>
<td>60</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Teacher Questionnaires</td>
<td>63</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Student Questionnaires</td>
<td>63</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Pilot and Field Testing</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Procedures: Autumn Quarter</td>
<td>64</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Observations</td>
<td>65</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Interviews</td>
<td>67</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Teacher Questionnaire</td>
<td>68</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Student Questionnaires</td>
<td>68</td>
</tr>
<tr>
<td>3.6</td>
<td>Procedures: Winter Quarter</td>
<td>69</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Observations</td>
<td>70</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Teacher Interview</td>
<td>70</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Student Questionnaires</td>
<td>71</td>
</tr>
<tr>
<td>3.7</td>
<td>Qualitative Data Analysis Procedures</td>
<td>72</td>
</tr>
<tr>
<td>3.8</td>
<td>Data Reporting</td>
<td>74</td>
</tr>
</tbody>
</table>

4. Results

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>76</td>
</tr>
<tr>
<td>4.2</td>
<td>Review of Testing Program in Spanish 104</td>
<td>76</td>
</tr>
<tr>
<td>4.3</td>
<td>Participant Profiles</td>
<td>79</td>
</tr>
<tr>
<td>4.4</td>
<td>Overview of Teacher Data Collected</td>
<td>82</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Classroom Observations: Autumn Quarter 2004, and Winter Quarter 2005</td>
<td>82</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Teacher Interviews</td>
<td>85</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Instructor Questionnaire and Responses</td>
<td>86</td>
</tr>
<tr>
<td>4.5</td>
<td>Organization</td>
<td>89</td>
</tr>
<tr>
<td>4.6</td>
<td>Anne</td>
<td>89</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Observations</td>
<td>89</td>
</tr>
</tbody>
</table>
4.6.2 Interviews
4.6.3 Anne’s Questionnaire Results
4.7 Beth
4.7.1 Observations
4.7.2 Interviews
4.7.3 Beth’s Questionnaire Results
4.7.4 Anne’s Questionnaire Results
4.8 Cathy
4.8.1 Observations
4.8.2 Interviews
4.8.3 Cathy’s Questionnaire Results
4.9 Diane
4.9.1 Observations
4.9.2 Interviews
4.9.3 Diane’s Questionnaire Results
4.10 Elena
4.10.1 Observations
4.10.2 Interviews
4.10.3 Elena’s Questionnaire Results
4.11 Summary of Observations, Interviews, Teacher Questionnaires
4.11.1 Summary of Behavior Patterns in Teacher Observations
4.11.2 Teacher Behaviors Exhibited Focusing on Grammar: Autumn and Winter Quarters
4.11.3 Washback Behaviors Focused on Textbook Reading and Writing Activities: Autumn and Winter Quarters
4.11.4 Teacher Behaviors and Possible Washback Focused on General Skill Building in Reading and Writing: Autumn and Winter Quarters
4.11.5 Teacher Behaviors From Past Teaching Experience and Testing Experience: Autumn and Winter Quarters
4.11.6 Overview of Interviews: Autumn and Winter Quarters
4.11.7 Themes: Autumn Quarter
4.11.8 Teacher’s Past Teaching Experience
4.11.9 The Role of Grammar in Language Learning
4.11.10 Instructor Preparedness and Familiarity with Content and Curriculum
4.11.11 Student Preparedness
4.11.12 Winter Quarter Interview
4.11.13 Theme: Winter Quarter
4.11.14 Familiarity with Course Content and Materials Related to the Test Resulting in Shifts in Instructional Emphasis
4.11.15 Student Expectations with Regard to Textbook Readings and the Tests
4.11.16 Summary of Interviews from Autumn and Winter Quarters
4.11.17 Teacher Questionnaire Overall Comparison
4.11.18 Summary of Two Highest Rated Responses
4.11.19 The Importance of Tests: Questions One, Two, Six, Seven and Eight ................................................................. 170
4.11.20 Questionnaire Summary ........................................................... 177
4.12 Student Questionnaires ........................................................................ 177
  4.12.1 Student Responses: Time Studied for each Chapter Test ............ 178
  4.12.2 Study Habits for Anne’s Students ............................................ 179
  4.12.3 Study Habits for Beth’s Students ............................................. 181
  4.12.4 Study Habits for Cathy’s Students .......................................... 183
  4.12.5 Study Habits for Diane’s Students .......................................... 185
  4.12.6 Study Habits for Elena’s Students .......................................... 187
  4.12.7 Summary of Student Reported Study Time ............................. 189
  4.12.8 Final Exam Student Responses: Perceptions of Tests and Instruction ........................................................................... 190
  4.12.9 Anne’s students ................................................................. 191
  4.12.10 Beth’s students ................................................................... 192
  4.12.11 Cathy’s students ............................................................... 193
  4.12.12 Diane’s students ............................................................... 194
  4.12.13 Elena’s students ............................................................... 195
4.13 Summary Students’ Final Exam Questionnaire ...................................... 196

5. Discussion .............................................................................................. 197
  5.1 Overview ......................................................................................... 197
    5.1.1 Washback ............................................................................... 197
    5.1.2 Spanish 104 Washback Study ................................................ 198
  5.2 Discussion of the Research Questions ............................................. 203
  5.3 Implications and Suggestions Arising From the Study ..................... 231
  5.4 Recommendations for Future Research ........................................... 236
  5.5 Conclusion ....................................................................................... 239

Bibliography ............................................................................................. 243

Appendices:

A . Class Observation Instrument ........................................................... 247

B . Teacher Instruments .......................................................................... 248
  Teacher interview ................................................................................ 248
  Second teacher interview ................................................................... 250
  Teacher questionnaire ........................................................................ 251
  Third teacher interview ...................................................................... 254
  Second quarter teacher interview .................................................... 256
<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Student Instruments</td>
<td>257</td>
</tr>
<tr>
<td>Student Instrument Tests 1, 2, 3</td>
<td>257</td>
</tr>
<tr>
<td>Student Instrument Final exam</td>
<td>258</td>
</tr>
<tr>
<td>D. Instructional Time Distribution (Teacher Classroom Observations)</td>
<td>261</td>
</tr>
<tr>
<td>Anne Autumn Quarter</td>
<td>261</td>
</tr>
<tr>
<td>Beth Autumn Quarter</td>
<td>263</td>
</tr>
<tr>
<td>Cathy Autumn Quarter</td>
<td>265</td>
</tr>
<tr>
<td>Diane Autumn Quarter</td>
<td>267</td>
</tr>
<tr>
<td>Elena Autumn Quarter</td>
<td>269</td>
</tr>
<tr>
<td>Anne Winter Quarter</td>
<td>271</td>
</tr>
<tr>
<td>Beth Winter Quarter</td>
<td>273</td>
</tr>
<tr>
<td>Cathy Winter Quarter</td>
<td>275</td>
</tr>
<tr>
<td>Diane Winter Quarter</td>
<td>277</td>
</tr>
<tr>
<td>Elena Winter Quarter</td>
<td>279</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>3.1</td>
<td>56</td>
</tr>
<tr>
<td>4.1</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>146</td>
</tr>
<tr>
<td>4.3</td>
<td>151</td>
</tr>
<tr>
<td>4.4</td>
<td>154</td>
</tr>
<tr>
<td>4.5</td>
<td>169</td>
</tr>
<tr>
<td>4.6</td>
<td>179</td>
</tr>
<tr>
<td>4.7</td>
<td>181</td>
</tr>
<tr>
<td>4.8</td>
<td>183</td>
</tr>
<tr>
<td>4.9</td>
<td>185</td>
</tr>
</tbody>
</table>
4.10 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarters for Elena’s students ................................................................. 187

4.11 Mean student response for each quarter for Anne’s students ......................... 191

4.12 Mean student response for each quarter for Beth’s students ............................. 192

4.13 Mean student response for each quarter for Cathy’s students ........................... 193

4.14 Mean student response for each quarter for Diane’s students ........................... 194

4.15 Mean student response for each quarter for Elena’s students ........................... 195

5.1 Identified washback behaviors, and the teachers who exhibited them most during the Autumn and Winter quarters ................................................................. 205

A.1 Instructional behaviors for Anne’s observed classes during Autumn 2004........ 261

A.2 Instructional behaviors for Beth’s observed classes during Autumn 2004........ 263

A.3 Instructional behaviors for Cathy’s observed classes during Autumn 2004........ 265

A.4 Instructional behaviors for Diane’s observed classes during Autumn 2004........ 267

A.5 Instructional behaviors for Elena’s observed classes during Autumn 2004....... 269

A.6 Instructional behaviors for Anne’s observed classes during Winter 2005......... 271

A.7 Instructional behaviors for Beth’s observed classes during Winter 2005......... 273

A.8 Instructional behaviors for Cathy’s observed classes during Winter 2005......... 275

A.9 Instructional behaviors for Diane’s observed classes during Winter 2005......... 277

A.10 Instructional behaviors for Elena’s observed classes during Winter 2005....... 279
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Total mean responses on questionnaire for Anne</td>
<td>100</td>
</tr>
<tr>
<td>4.2</td>
<td>Total mean responses on questionnaire for Beth</td>
<td>110</td>
</tr>
<tr>
<td>4.3</td>
<td>Total mean responses on questionnaire for Cathy</td>
<td>122</td>
</tr>
<tr>
<td>4.4</td>
<td>Total mean responses on questionnaire for Diane</td>
<td>131</td>
</tr>
<tr>
<td>4.5</td>
<td>Total mean responses on questionnaire for Elena</td>
<td>142</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Testing in Public Education

The American public has cultivated an obsession with testing in education, largely as a result of intensive educational reform. The national curricular standards (Goals 2000), signed into law by President Clinton in 1994, created monumental changes in education at all levels. Although the national curricular standards are voluntary, there is increasing pressure to conform, as state and local standards become established and accepted (Schwartz, 2002). The changes brought about by the standards were perpetuated with the No Child Left Behind act signed into law in 2002 by President Bush. On the U.S. Department of Education’s website it is possible to find a plethora of information for parents, teachers, and the general public, in Spanish and in English about this monumental educational initiative. According to information on the website, with passage of No Child Left Behind, Congress reauthorized the Elementary and Secondary Education Act (ESEA)--the principal federal law affecting education from kindergarten through high school. In amending ESEA, the new law represents a sweeping overhaul of federal efforts to support elementary and secondary education in the United States. It is built on four common-sense pillars: accountability for results; an emphasis on doing what works based on scientific research; expanded parental options; and expanded local control and flexibility” (http://www.ed.gov/nclb/landing.jhtml?src=pb, retrieved on August 24, 2004).
Consequently, the laws governing curriculum, and how to assess that curriculum, were redistributed and in some cases, reassigned completely. Whereas education was once the purview of state and local governments, accountability in education has become everyone’s responsibility. The most common way to establish accountability has become standardized testing.

The national curricular goals for Education presented to the American public incorporated standards for the subject areas of math, science, history, and English, with the intent of providing “high expectations for all learners” (Tucker and Codding, 1998, cited in Schwartz, 2002). These standards serve as benchmarks by which all students are measured at particular levels. To determine whether or not learners have met these standards, students are evaluated by mandated tests in the various subject areas. Successfully meeting these standards allows students to progress through the academic system. However, such tests can serve other purposes as well. Currently in Virginia, for example, the results from these mandatory tests are increasingly being used to determine whether schools are operating properly… Schools that meet standards are being given waivers from state regulations, paperwork requirements, and required instructional time. Schools that do not meet standards often have to accept help from teachers in schools with higher test scores (master teachers), whose function is to spur them on to implement complex improvement plans and new instructional methods that low-scoring schools are required to adopt (Cunningham & Sanzo, 2002, p. 63).

Testing is assuming a dominant role in American public education, and the term “high stakes tests” is more relevant and telling than ever before.

1.2 Foreign Languages

Not included in the initial list of previously mentioned subject areas to be tested was the area of foreign languages. Only through the results of massive lobbying efforts
on the part of numerous national foreign language associations, was foreign languages finally accepted in 1993, as the seventh and final curriculum area to be included in the Goals 2000 project. The language associations included but were not limited to the American Council on the Teaching of Foreign Languages (ACTFL), the National Committee on Languages (NCL), the American Associations of Teachers of French (AATF), Spanish and Portuguese (AATSP), and German (AATG) (Schwartz, 2002).

Upon the acceptance of foreign languages as an integral part of public education in the United States, the national Standards for Foreign Language Learning, commonly referred to as the 5Cs, were completed in 1995 through collaboration among the previously mentioned organizations: ACTL, AATF, AATG, and AATSP. These content standards provide goal areas that serve as guides for the language profession as educators lead language learners through language development. Although the Standards for Foreign Language Learning do provide sample progress indicators throughout the document as possible means of assessing students’ progress, they do not provide specific skills and abilities, as do the ACTFL Proficiency Guidelines. Nor does the document seek to provide a specific curriculum, or recommend scope and sequence for a language program (Standards for Foreign Language Learning, 1996, p. 13). The 5Cs—communication, cultures, connections, comparisons, and communities—are meant to provide a holistic framework, and lend themselves well to a communicative approach to language learning. Their holistic nature is reflected in the fact that the competencies in one area are fundamentally linked to competencies in each of the areas (Schwartz, 2002), and they support a communicative approach to language learning, as the learner is expected to use the language for a variety of purposes, and in a range of modes. These
communication modes consist of interpersonal, interpretive, and presentational where each mode focuses on a particular link between the language and the underlying culture (Standards for Foreign Language Learning, 1996, p. 32). The application of the Standards as a guide for language learning is one of the reasons that significant and dramatic changes have occurred in the way that teachers instruct and assess language skills.

Another reason for the changing foreign language environment is the advent of communicative language teaching (CLT), which posits that functional language ability is developed through a learner’s participation in events and activities demanding communication. As Savignon (2002) explains, “learners’ communicative needs provide a framework for elaborating program goals with regard to functional competence. Functional goals imply global, qualitative evaluation of learner achievement as opposed to quantitative assessment of discrete linguistic features” (p. 4). And yet, the shift to standardized tests has made quantitative assessment the dominant paradigm in the evaluation of student learning, resulting in a mismatch between what students learn to do in the communicative classroom and what standardized tests measure. Thus, the combination of the ‘communicative revolution’ and the adoption of the national Standards for Foreign Languages has created a controversy in terms of how best to assess learners’ abilities and competencies, especially within the current atmosphere of highly public accountability. Although foreign languages are not yet included in state-wide proficiency tests in many states, that time may not be far off. Indeed, a move toward national testing of students’ abilities in Spanish is now underway. Furthermore, the Foreign Language National Assessment of Educational Progress (FL NAEP) has been
in development, and was originally scheduled to be piloted in 2003. However, the scheduled implementation has since been revised and is currently on hold. Although there are many more hurdles to clear before foreign languages may be subjected to the same tests as the other core subjects, such testing appears to be inevitable.

1.3 Washback

As illustrated in the earlier example regarding the Virginia school systems, tests are frequently used to spur educational change and innovation, in addition to establishing accountability. This practice results in a phenomenon referred to as “washback,” a term used to describe the impact (negative or positive) that testing has on various aspects of a teaching program, including teachers, classrooms, and students. According to one of the founding washback researchers, J. Charles Alderson:

We now know, for instance, that tests will have more impact on the content of teaching and the materials that are used than they will on the teacher’s methodology. We know that different teachers will teach to a particular test in very different ways. We know that some teachers will teach to very different tests in very similar ways. We know that high-stakes tests—tests that have important consequences for individuals and institutions—will have more impact than low-stakes tests, since what is a trivial consequence for one person may be an important matter for another (p. ix-x, 2004).

1.4 Washback in Foreign Languages

Although “washback is a phenomenon that is of inherent interest to teachers, researchers, program coordinators/directors, policy-makers, and others in their day-to-day educational activities” (Cheng, Watanabe, & Curtis, 2004, p. xiii), and research on washback in the area of foreign languages is increasing, there exists a need for a much deeper understanding of the nature and extent of washback, especially as state mandated subject area tests are imposed upon school districts throughout the United States. After
foreign languages become a part of this new standardized testing culture, and foreign language teachers face growing pressure to “teach to the test,” for example, washback will be a phenomenon of extreme importance in the field of foreign language pedagogy.

Foreign language specialists have been working to achieve the kind of subject matter status that warrants standardized testing, and to become part of the core curriculum of subjects tested. Moreover the national *Standards for Foreign Language Learning* indicate that it is necessary to relate foreign language education to other disciplines so that students are able to further their knowledge through the foreign language. In other words, the foreign language profession needs to scrutinize the development, effect, and long standing implications of the most common assessment and evaluation practices for the discipline. In this regard, Savignon (2002) points out, “controversy over appropriate language testing persists, and many a curricular innovation has been undone by failure to make corresponding changes in evaluation” (p. 4). With the growing emphasis on high stakes testing in education today, there appears to be an accompanying belief in the need to use tests to create systemic validity, meaning that tests are integrated into an educational system to try to demonstrate that the introduction of such tests improves learning (Shohamy, 2001). Although there is controversy as to whether testing actually improves learning, several studies have shown that it does affect teaching, and therefore must ultimately relate to learning in some context (Alderson & Hamp-Lyons, 1996; Cheng, 2002; Shohamy, Donitsa-Schmidt, & Ferman, 1996; Watanabe, 1996). At present, the picture of how testing relates to teaching remains unclear. The teacher behaviors that occur as a result of language tests and the impact on instruction, i.e., washback, are the primary focus of this study. Through observations,
teacher interviews, teacher questionnaires, and student questionnaires, this study describes washback as it was observed in the intermediate Spanish foreign language classroom. By providing a detailed and thorough description of washback behaviors exhibited by teachers, the foreign language profession can utilize this information to improve instruction, testing practices, professional development programs, teacher training programs, and ultimately, affect student achievement.

For the purposes of this study, the definition of washback offered by Cheng, Watanabe, and Curtis (2004) in the preface to their book *Washback in Language Testing*, served as the foundation. They state, “washback…refers to the influence of language testing on teaching and learning” (p. xiii). Although many studies have investigated washback to determine a) whether it actually exists, and b) the nature and extent of its effects, these studies have focused predominantly on the use of high stakes tests, that is, those used to make decisions that affect students’ admission into schools, placement in courses, and post-school directions in life. Research studies have also examined the use of tests as a tool for initiating curricular change (Alderson & Hamp-Lyons, 1996; Cheng, 2002; Shohamy, 2003). In contrast, this study focuses on the impact of a low stakes testing program that was an existing part of the intermediate Spanish language curriculum, and not implemented for the purpose of affecting any changes.

There remains a need for more empirical studies of washback and a deeper understanding of the complexities that surround it, including the descriptive factors that reveal why “tests will have different amounts and types of washback on some teachers and learners than on other teachers and learners” (Alderson & Hamp-Lyons, 1995, p.
296). McNamara described the complicated, and sometimes seemingly contradictory nature of washback, as follows:

It has been claimed that communicative, performance-based tests would have beneficial washback on the teaching that led up to them (cf. Wesche, 1987), as the teaching would focus on preparing students for the representative communicative tasks, in the tests, and thus indirectly for the real-world communicative tasks beyond the test. The assumed negative washback of non-performance based tests was held to be an unfortunate constraint on curriculum reform (1996, p.23).

However, it has been shown that the picture is not that simple and clear cut; simply including tasks or information on a test does not guarantee that language learners will learn material, or be better prepared for real world use of the language. Nor does it ensure that teachers will teach the material effectively. Previous studies conducted on washback have focused on different means of measuring it, and identifying factors that contributed to its effects. In the earliest studies, it was common for researchers to use test results and self-generated accounts of classroom occurrences from teachers and students. In 1993, Wall and Alderson conducted a study that also included classroom observation as a data gathering technique. Presently, classroom observations, interviews, and questionnaires are standard procedures in qualitative research focusing on washback. In this study, classroom observations, teacher interviews, and teacher and student questionnaires, were used to clarify imprecise factors that may be linked to washback (Alderson and Hamp-Lyons, 1996), as well as provide a detailed description of washback in the intermediate Spanish foreign language classroom.

1.5 Statement of the Problem

The adoption of communicative language teaching (CLT) as the dominant approach in foreign language classes, and the publication of the national Standards for
Foreign Language Learning (1996), have changed the way foreign language teachers are expected to evaluate their students’ skills and overall competency. Research has indicated that there are certain evaluation instruments or activities that influence the classroom teaching of certain teachers in certain situations. However, the descriptive characteristics of these tasks, teachers, and situations are not yet clearly defined or articulated. Additionally, well known washback studies have repeatedly indicated that although tests may directly affect the content that is taught to the students, tests do not necessarily affect the ways that teachers teach (Cheng, 2004; Fermin, 2004; Wall & Alderson, 1996). To gain further understanding of how the evaluation process affects classroom instruction, is affected by other factors, and may affect student achievement, it is necessary to conduct further research on test washback. As Wall emphasizes,

> despite increasing interest in washback and other types of impact, there is still not enough research taking place to inform testers of the kinds of influence that it is reasonable to expect given certain test characteristics and certain features of the educational context, nor is there a comprehensive enough model to help educators avoid the negative effects … There is a clear need for more research into the nature of test washback and other forms of impact, and for a model which will guide testers on how to best produce and introduce tests in order to influence teaching (Wall, 1997, p. 298).

Furthermore, the majority of the studies on washback have focused on contexts that include the use of high stakes tests, where the language of the test has been English as a second or foreign language. The studies have been conducted most commonly in educational settings outside of the United States, within an educational context that requires significant foreign language study for all students. Despite its popularity in foreign language programs in America at various school levels (elementary school through university), washback in the Spanish foreign language classroom has not been
significantly researched, nor has it been researched in a low stakes testing environment. Therefore, a major void in our understanding of washback in foreign language pedagogy exists.

This study addressed that void by researching testing and washback in the low stakes testing environment of an intermediate foreign language Spanish course within a large undergraduate Spanish language program at The Ohio State University. The program was selected for several reasons. These reasons included but were not limited to:

- The program had provided the same amount and type of training for all of the teachers who were about to embark on teaching the Spanish 104 course for the first time.
- The majority of the students enrolling in the Spanish 104 course had the same amount of experience within Ohio State’s Spanish language program. Therefore, they were equally familiar with the programs’ objectives and requirements.
- The Spanish language program used standardized chapter tests that each and every student enrolled in the Spanish 104 course experienced.

Thus, there was a significant degree of commonality among both the teachers and the students in the intermediate program, thereby ensuring a consistent research base within which to operate. Furthermore, because Spanish is the most commonly taught foreign language in the United States, and the transition from elementary to intermediate level foreign language courses can be problematic for both students and teachers as a result of the shift in emphasis on particular skills, this setting provided an ideal opportunity to investigate relationships between testing and teaching, and how they may affect student achievement. Research at the intermediate university level was also advantageous
because there are many commonalities between intermediate programs at the high school level and the university level, thus extending the possible application of the results. Although the results of this study are directly pertinent only to the university level of teaching, the information gained is also of importance to all intermediate level Spanish language teachers, especially as the standardized testing movement gains strength in the public school system. It should also be noted that conducting this study at the university level was not only feasible, but also desirable given the university quarter schedule, and the large number of sections of the course that are taught each quarter. Both conditions allowed for increased opportunities to gather data.

1.6 Significance of the Study

The importance of learning more about washback is threefold. First of all, test developers, especially those writing tests for a large university undergraduate Spanish language program like the one described in this study, must be experts in the field of foreign language assessment. They must have an understanding of important issues in test development such as validity and reliability (Messick, 1996), competencies and skills to be tested (Bachman, 1985), and how test takers interpret test items (Oller, 1991), as well as how each of these aspects may relate to the concept of washback. A thorough understanding of these aspects will allow test writers to use the knowledge and information as it relates to washback in order to develop tests designed to assess the goals and objectives of the language program, and promote second language acquisition. This study provides some of the knowledge and understanding needed to help future teachers and Spanish language test developers to design instruction and tests that properly account for washback.
Second, the intermediate Spanish language classroom teacher has a vested interest in knowing how washback affects instruction and assessment, and how to best recognize the effects of washback, either positive or negative, on her/his teaching. With its emphasis on a group of such teachers, this study offers insight directly relevant to their needs.

Third, learning more about washback in this environment has important implications for teacher education programs that are designed to train new instructors, to introduce valuable classroom techniques and strategies, and to review important issues in the development of course syllabi and teaching materials. If, as earlier studies indicate, testing does not necessarily result in differences in teaching, and it is assumed that teaching should respond to changes in testing practices, then teacher training programs need to prepare new teachers to be aware of, and to be sensitive to, relationships between testing and teaching. This research provides valuable information to assist administrators and program directors in providing the necessary professional development and guidance to teachers teaching and testing foreign language skills.

When foreign languages become part of the American national testing program, as reported by the National Assessment of Educational Progress (NAEP), then all foreign language educators should be as informed as possible about the various aspects of the testing process as they relate to classroom instruction and learning. It is important to seek to understand how teacher behavior is reflective of tests, and how that behavior can be influenced in a positive manner, so that students can meet the academic benchmarks that are determined for them. In an era in which standardized testing has become a dominant paradigm in the assessment of learning, there is an increased need to understand
washback, particularly because the higher the stakes in testing (and with standardized testing, the stakes are normally high), the better the conditions become for the occurrence of washback. The results of this study facilitate the endeavor to better understand washback as it relates to foreign language teaching and testing.

1.7 Research Questions

The study describes washback in the intermediate Spanish foreign language classroom within a large university program. The research questions provided the framework necessary to uncover the descriptive characteristics of teacher behaviors that relate to washback. These questions were as follows:

1. In a low stakes testing context, what form(s) does washback take? How is washback manifested?

2. What characteristics and factors draw distinctions between washback and teacher behaviors that reflect adjustments to the course as opposed to test related washback?

3. What similarities and differences exist with respect to teacher behaviors and washback based on teacher perceptions, student perceptions, and the researcher’s perception?

4. What elements of an intermediate testing program encourage positive washback?

5. What elements of an intermediate testing program encourage negative washback?

6. Which of the student behaviors previously reported in the literature as indicating washback, were observed in the classroom? What additional behaviors suggesting washback were observed?
1.8 **Methodology**

This qualitative study was conducted in intermediate Spanish language classes at The Ohio State University during the Autumn 2004 and Winter 2005 quarters. The classes that served as the primary focus of the study were sections of Spanish 104 taught by experienced TAs and Lecturers in the Department of Spanish and Portuguese. Each teacher was teaching the course for the first time in the Autumn 2004 quarter. The reason for selecting subjects who were teaching Spanish 104 for the first time is that they had limited knowledge of the testing program. The fact that the testing program for the elementary language program and the intermediate language program are different was important because each emphasized distinct material, and placed a different focus on language skills. Therefore, this provided a unique opportunity to observe changes in the teacher behaviors from the Autumn to the Winter quarters as the teachers’ knowledge of the testing program increased.

The decision to select teachers with some experience came from the assumption that confounding factors affecting teacher behaviors, such as confidence, inexperience, limited repertoire of teaching strategies, limited exposure to a variety of students, learning styles, and behaviors, and issues of classroom management could be minimized with this group. Although all of the teachers had taught at least one year in the Spanish language program at The Ohio State University, some were relatively new to teaching.

Selecting specific subjects is argued by Patton (1987) as presented in Watanabe (2004): “the selection is not to be made at random, but purposefully” by “selecting the *information-rich cases* for study in depth” (p. 52). Data gathering techniques included classroom observations, teacher interviews, and teacher and student questionnaires. The
class observations, teacher interviews and teacher questionnaire, provided details for the
description of washback, and data relating to the forms and characteristics of washback.
The student questionnaires elicited information about potential student behaviors
indicative of washback, and students’ perceptions of their teacher’s behavior with respect
to tests and instruction.

1.9 Definition of Key Terms

**authentic assessments**: evaluations which seek to mimic real world activities as closely
as possible in order to provide a more realistic context in which to assess the skills being
measured

**background washback**: washback behaviors that occur in anticipation of a test that may
include but are not limited to heightened anxiety, frequent questions about the format and
content of a test, focus on procedures for test; may be exhibited by teachers or students

**contextualized assessments**: evaluations that provide a context for the activity or task in
order to relate it to the learner’s environment, to facilitate the understanding of the nature
of the activity and increase the degree of authenticity

**focused washback**: washback that occurs when there is direct knowledge of a specific
test

**intermediate language class**: Spanish 104 (Intermediate Spanish II), which has the
following course description: reading of Spanish short stories, plays, and novels with
attention to literary appreciation; development of basic language skills; course conducted
in Spanish

**negative washback**: activities incorporated into classroom instruction as a result of the
test that inhibit, or do not promote the development of, language acquisition and/or the
development of language skills (reading, writing, speaking, listening) e.g., a conscious focus on discrete point grammatical activities and avoidance of skill building activities, in order to prepare students for a specific portion of the test

**non test impacting observation:** observation of a class session that focused on material, or stressed activities that were not directly included on the chapter test

**positive washback:** activities incorporated into classroom instruction as a result of the test that promote the development of language acquisition, and/or the development of language skills (reading, writing, speaking, listening) e.g., a conscious focus on task-based activities that mimic real world situations where students are required to use the target language to successfully complete a task. These include, but are not limited to, writing a letter, gathering information from an authentic text, listening to and comprehending authentic language

**student behaviors indicative of washback:** activities and strategies that students use to prepare for a test as a result of class activities and test knowledge, and incorporate into their repertoire of language learning strategies

**task-based assessments:** evaluations in which students must use their skills in a particular area of the language (reading, writing, speaking, or listening) to complete a task

**teacher behaviors indicative of washback:** instructional methods and strategies that become classroom practice as a result of the format of the test, and are incorporated into the classroom teaching repertoire (e.g., these may include, but are not limited to, introducing materials not included in the text, eliminating or reducing the use of materials not included in the test, focused class time devoted to reading activities, focused class
time devoted to writing activities, reduced class time for speaking activities, focused class
time on direct test like activities)

**test impacting observation:** observation of a class session that focused on material, or
stressed activities that were directly included on the chapter exam

**washback:** the influence of language testing on teaching and learning (Cheng,
Watanabe, Curtis, 2004)

1.10 Assumptions of the Study

1. Washback is observable and a researcher is able to discern between positive and
   negative washback.

2. Selecting teachers with some previous teaching experience would reduce the effects
   of confounding factors such as confidence, inexperience, a limited repertoire, limited
   exposure to a variety of students, learning styles, and behaviors, and issues of
   classroom management that influence teaching and the observance of washback.

3. Teachers teaching in the intermediate program for the first time would have limited
   knowledge of the intermediate testing program.

4. Students enrolled in the intermediate language program would have limited
   knowledge of the intermediate testing program.

5. Intermediate language classes would exhibit observable signs of washback due to
   the tests used in the program.

6. The observations, interviews, and questionnaires used in gathering data were valid
   and reliable for the purpose of the study, as a result of reliability checks, pilot testing,
   and field testing.

7. Teachers new to the intermediate program, i.e., teaching Spanish 104 for the first
time, and unfamiliar with the exact testing process would alter their teaching after the first test (e.g., they would incorporate more reading and writing activities into the class sessions, and homework, as each test contains writing activities based on authentic readings from the text, valued between 40-60 percent of each test).

8. Teachers would adjust their teaching to a greater degree to reflect knowledge of the testing program during the second quarter of teaching Spanish 104, i.e., exhibiting more signs of focused washback.

9. Classes chosen for observations and determined as test impacting and non-test impacting would be based on the pre-determined class calendar published prior to the start of the quarter.

1.11 Limitations of the Study

1. Only intermediate college level Spanish students were the focus of the study. Thus results are only generalizable to the extent that they describe this population.

2. Classroom observations may have had an effect on teacher performance and behavior.

3. Classroom observations may have had an effect on student performance and behavior.

4. Although classroom observations were made based on the class calendar, instructors are not required to follow the calendar exactly as it is detailed. This resulted in some observations that did not focus on the lesson from the calendar, especially during the Autumn quarter. Consequently, the lesson observed may or may not have been the test impacting lesson originally determined by the researcher.
5. The researcher’s participant role in the research (during class observations) may have had an effect on teacher behavior and student behavior.

6. Background washback may have diluted the observations of focused washback.

7. For some teacher behaviors, it was difficult to distinguish between quantity due to washback, and the quantity due to increased familiarity with course content and experience.

8. Some teacher participants in the study did not have significant teaching experience; therefore, a limited teaching repertoire affected washback behaviors.
2.1 Historical Perspective

David Patrick Barnwell writes in the preface to his book *A History of Foreign Language Testing in the United States*:

unlike other fields of human endeavor, foreign language pedagogy preserves but a superficial memory of its own history and benefits little from the patrimony of the great figures of its past. The ideas and labors of those who dominated the profession decades ago are forgotten, and the contribution they made have been erased from memory (1996, p. vii).

In 1961, Robert Lado published the seminal text, *Language Testing: The Construction and Use of Foreign Language Tests, A Teacher’s Book*. Presumably, in view of the numerous changes that have occurred in the discipline of foreign language teaching, nothing from this text would apply to language teaching and testing as they exist today. Surprisingly, however, there are some underlying principles that are quite pertinent and applicable to the present state of affairs with regard to foreign language teaching and testing. As Lado observed in this book on language testing,

Teaching methods have been largely the collected practice and teaching habits of particular teachers who reacted to the memory of their own experience in learning, the example of their own teachers, and the fashion which seemed prevalent or attractive at the particular moment in the history of language teaching (1961, p. 17).
That is, the practices of one generation of teachers are at least in part a reflection of how they were taught. In this case, foreign language teaching today is not only a product of currently existing conditions, but also contemporary practices having roots in the pedagogy of the past. Just as what teachers do in the classroom is very much affected by what their prior classroom experiences consisted of, their preconceptions regarding teaching formed as a result of these experiences (Crandall, 2000; Freeman & Johnson, 1996). Likewise, testing and assessment in the foreign language classroom do not appear to be immune to the influence of prior experience (Barrette, 2004). Decades later, current assessment practices reflect to some extent what Lado believed. Although foreign language pedagogy has moved through the grammar translation method, the audio lingual method, and earnestly embraced the communicative language teaching approach, in some ways language testing has met and withstood the challenges posed by these changes in instructional practices, and in other ways, it has consciously sought to avoid them. Instruction is inevitably linked to testing because what is taught in the classroom most often becomes the focus of tests and assessments. Therefore, it is virtually impossible to consider one without the other, and the influence that each has on the other.

History has demonstrated that, whichever teaching method dominates, what and how students learn will affect what and how teachers test. For example, during the 19th century, American colleges routinely administered tests of foreign language ability to incoming students. French was the language of the test, as this was the most popular foreign language studied. The most common testing approach was to require translation of sentences and longer passages, together with some more direct tests of grammatical knowledge. As Barnwell (1996) explains, the translations were usually of highly
artificial and/or nonsensical passages or sentences, and the grammar exercises even included questions that “at the intermediate level…tested the ability to conceptualize and articulate metalinguistic knowledge, such as: ‘How do the modal auxiliaries differ from ordinary weak verbs, and how from strong verbs?’” (p. 3). In other words, tests were aimed at revealing the amount of information learners possessed about the target language. At the same time, says Barnwell, the focus of language teaching, like testing, was to have students know about the language, but not to use it (p. 3). Surprisingly enough, although the goals of language instruction have shifted greatly, and the languages taught have expanded well beyond French, there are still institutions that use this type of test today. Hence, teaching practices have long mirrored the kind of knowledge required on tests, an indication of early washback.

Historically, social and political changes have influenced the enrollment patterns in foreign language classrooms, as is observable with present enrollments in Spanish. Ironically, when the College Board administered its first language exams in 1901, the only languages tested were French, German, and Latin. Spanish was not offered until the following year, and a mere nine students participated in the exam (Barnwell, 1996), a bold contrast to the over 76,000 Advanced Placement Spanish Language and Literature exams administered in May 2004. During World War I, changes occurred not only in the student enrollments in language classes (e.g., German declined substantially), but also in the desire to delve deeper into the field of language testing (Barnwell, 1996).

As foreign language testing became more defined as a discipline, new ways of thinking about measuring language skills emerged. Standardized language testing first
appeared in 1919 with the publication of Charles Handschin’s *Silent Reading Test in French and Spanish*. Handschin, a faculty member at Miami University in Ohio, developed a test consisting of a paragraph in the target language followed by comprehension questions, and a grammar-based test. An important consequence of this test was the fact that Handschin developed basic principles to be applied when writing a language test. Although this early approach to the testing of language may seem primitive and lack insightful information about what a language learner actually knows, as Barnwell states,

> in the early 1920s those involved in teaching for the first time were consciously reflecting on the question of how to assess achievement and in doing so, were actually creating the very idea of foreign language testing as an academic activity (1996, p. 20).

Testing through the decades of the 1920s, 30s, 40s, and 50s shifted frequently and dramatically as a result of further changes in the political and social climate in the United States, and also as a result of inquiries into the principles of foreign language testing. Statistics were eventually used to measure the efficacy of tests, and the range of skills and abilities intending to be assessed was enlarged, thus bringing such crucial aspects of testing as reliability and validity into play. Tests began to be used regularly for placement purposes, and changes were made to established tests, such as those offered by the College Board, in order to include different skills and types of abilities. Another example of tests being used to assess different abilities was a test developed by James Tharp of The Ohio State University. He developed a test about French civilization using a matching format to test cultural knowledge.
The development of the Army Specialized Training Program (ASTP) during World War II had a significant and continuing influence on language teaching, training, and testing. The greatest influence that can be extracted from this phase of foreign language test development is the emphasis that ASTP placed on oral language. As a result of army personnel needing communication skills in specific foreign languages, and their need to understand and be able to use colloquialisms in order to communicate in the field, trainees were instructed in skill building using what came to be known as the Audiolingual Method of instruction. It was accepted that language learners needed to be able to communicate with native speakers via oral and auditory communication. The ALM used colloquial, sociolinguistically appropriate language in the dialogues and recombination narratives, …a feature that was missing in older methods such as grammar translation. The focus on oral skills led to good pronunciation and accurate speech, …[and] methods…also stressed the teaching of culture [to] prepare students to deal in some measure with everyday situations in the target language community (Omaggio Hadley, 2001, p. 112).

Although assessment methods continued to focus largely on grammatically based activities, the trainees were expected to be able to use the language as well. The development of the military language programs impacted the language programs offered in schools, as seen in the shift to the Audiolingual methodology (further described in the next section). This method was a major part of foreign language instruction in the second half of the 20th century, and continues to be practiced today. The influence of the initial underlying principles of language instruction focused on language use on contemporary language instruction can be linked to the ideology of the ASTP many years ago (Barnwell, 1996).
2.2 Language Teaching and Test Design

As a result of the various influences on foreign language instructional programs, approaches to foreign language instruction have been quite varied. The early approaches to language teaching held that the instructor was the central figure in the classroom. As the foreign language profession moved from grammar and text-translation methods to a more “oral” approach, the instructor-as-authority-and-expert was codified in a teaching method called Audiolingual Methodology, commonly referred to as ALM or audiolingualism. ALM was predicated on the marriage of behaviorist psychology and then-current structural linguistics. According to behaviorist psychology, all learning—verbal and nonverbal—takes place through the process of habit formation. Habits are formed through repetition, imitation, and reinforcement. In ALM, language habits were formed by memorizing dialogues and practicing sentence patterns, usually through drills that required learners to imitate and repeat what their instructors said. Second language acquisition (SLA) was seen to be the replacement of first language habits by second language habits (Lee & VanPatten, 2003, p. 9).

This method did not require that students understand the language they were using, and student output of the language was restricted to the expected responses. Consequently, language instructors who at first were enamored with this method, became disenchanted:

The methodology had not delivered what it had promised: bilingual speakers at the end of instruction. [And]…the method did not take into account the variety of learning styles and preferences of students, favoring those who learned best “by ear” rather than visual learners and ignoring the needs of learners who wanted to understand the rules of the new language and its grammatical system (Omaggio Hadley, 2001, p. 112).

As methods of teaching have changed, so has the role of the instructor. This has been particularly notable during the “communicative revolution” that has transformed foreign language teaching since the early 1980s. Communicative language teaching (CLT) is a method that seeks to provide “learners… with opportunities to express real information
and not merely the information in drills” (Lee & Van Patten, 2003, p. 22). Hence, the role of the teacher has become facilitator of communication rather than the dominant force in the classroom. Rather than teachers, learners and their communicative needs are at the center of instruction, and teachers must provide opportunities for students to engage in real language use so as to express and interpret authentic messages in the target language (Lee & VanPatten, 2003). As a result of this, assessing the skills that are practiced and acquired in language classrooms today has become a much more complex and demanding task than in the past. Language testers now face the challenge of measuring students’ real-life communicative ability instead of simply their knowledge of the grammar and vocabulary of the target language.

2.2.1 Testing ‘Communicative Ability’

‘Communicative ability’ is a difficult construct to define and a harder construct to test because it cannot be tested completely via an objective format test:

…test construct refers to those aspects of knowledge or skill possessed by the candidate which are being measured. …Defining the test construct involves being clear about what knowledge of language consists of, and how that knowledge is deployed in actual performance (language use). Understanding what view the test takes of language use in the criterion is necessary for determining the link between test and criterion in performance testing. This is not just an academic matter. It has important practical implications, because according to what view the test takes, the ‘look’ of the test will be different, reporting of scores will change, and test performance will be interpreted differently. The difference of format between paper-and-pencil tests and performance tests is not just incidental; it reflects an implicit difference between views of language and language use (McNamara, 2000, p. 13).

Much of the process can be subjective, highly personal, and influenced by programmatic needs. As teachers seek to determine the level of communication that a student can perform at, increasingly students are being asked to communicate in situations mimicking
real world scenarios re-created for the classroom, in role playing situations, or even in
dramatic representations. Therefore, designing a test that measures this is a challenge for
test writers and teachers.

Performance tests are designed to test communicative competence. However, tests are complicated by the specific components that define communicative competence. Canale and Swain described aspects of competence that included grammatical, sociolinguistic, strategic, and discourse competence (McNamara, 2001, p. 18). Since the 1980s, when these concepts were first introduced, to the present, the concepts have been reformulated and further developed. Additionally, the factors other than knowledge that are part of performance testing, such as “…confidence, motivation, emotional states, the identities of candidate and interlocutor” have been emphasized as a means to understand what role these play in communication (McNamara, 2001, p. 19). Whereas the specific emphasis of these competencies varies according to the test task, and because each test may have a different and specific purpose, it becomes difficult to produce tests that accurately and fairly measure a candidate’s abilities in a particular set of circumstances.

2.2.2 Learning Theory

In measuring a student’s abilities, it is also critical to take into account how students learn, and a teacher’s beliefs regarding how this process occurs. For example, Cognitive Learning Theory

…views the learner as one who acts, constructs, and plans rather than simply receives stimuli from the environment. Therefore, a complete understanding of human cognition would require an analysis of strategies used for thinking, understanding, remembering, and producing language (McLaughlin (1990) in Omaggio Hadley, 2004, p. 65)
Depending upon how learners are able to organize new information internally for processing and learning, and how teachers organize instruction to enhance learning, affects the process of meaningful learning. In the classroom, meaningful learning is described as “relatable to what one already knows and thus can be easily integrated into one’s existing cognitive structure (Omaggio-Hadley, 2004, p. 68).” The task that a learner is required to perform on an assessment will demand different degrees of attention from the learner, different processing skills, and even result in different levels of production. Therefore, teacher beliefs regarding how students learn is an essential consideration when planning instruction and developing tasks to determine students’ level of abilities (Omaggio-Hadley, 2004).

Nevertheless, as the profession is committed to communicative language teaching, it also recognizes that the traditional paper-and-pencil tests will not accurately depict a candidate’s knowledge and abilities. A candidate can demonstrate written communication on paper using her/his reading and writing skills; however, oral communication and auditory comprehension need to be demonstrated in ways that demand students use those skills. Therefore, albeit slowly, language testing has moved forward with the development of tests and assessments that intend to find out not so much what the learner does not know, but rather what the learner does know, and how well s/he can use that knowledge in real-life communication.

2.3 Influences on Testing in the Communicative Foreign Language Classroom

The aforementioned changes in foreign language education that have occurred during the past decades have resulted in obvious changes in classroom assessment practices. Two of the most recent prominent conduits for change have been the national
Standards for Foreign Language Learning, and the adaptation of communicative language teaching (CLT) as the preferred method of instruction. The Standards for Foreign Language Learning articulate clear expectations and outcomes for what students should be able to do with their language skills as a result of successful language study. These expectations and outcomes are reflected in what are called the 5Cs of the Standards and include Communication, Cultures, Connections, Comparisons, and Communities. Briefly these encompass the following outcomes:

- students will learn to communicate in languages other than English (Communication),
- students will gain knowledge and understanding of other cultures (Cultures),
- students will connect with other disciplines and acquire information (Connections),
- students will develop insight into the nature of language and culture (Comparisons),
- and students will participate in multilingual communities at home and around the world (Communities) (Standards for Foreign Language Learning, 1996, p. 9).

Through collaboration with eleven different professional organizations, language standards were developed. As a result of this collaboration, teacher education standards were also developed, as well as state standards for foreign language learning and curricular frameworks that mimic the national standards. Because “…the standards are grounded in a combination of solid research, strong and sequential curriculum, and effective instructional practices”, much has been accomplished in reaching a strong national consensus (Philips, 1999, p. 2):

Standards broadened the content range of language learning by venturing well beyond the traditional four skills of listening, speaking, reading, and writing and the occasional study of culture. … The representation of all five goals as interlocking circles signifies that all should be systematically incorporated into language instruction at all levels (Philips, 1999, p. 3).

While the Standards have been developed in order to focus and strengthen foreign language teaching and learning, it is also anticipated that the National Assessment of
Educational Progress (NAEP) will incorporate these goals into assessment tools to be used at the national level in the United States. The long awaited large-scale standardized test of foreign languages, including Spanish, has yet to be developed; however, it has been in the planning and pilot stages for some time:

The inaugural NAEP Foreign Language assessment is currently under development. The assessment is designed to assess twelfth-grade students, who have learned Spanish in a variety of ways and for different lengths of time, at the national level only. A pilot test was conducted in the fall of 2003. On March 6, 2004, the National Assessment Governing Board postponed the planned 2004 administration (http://nces.ed.gov/nationsreportcard/foreignlang/, retrieved February 14, 2005).

In order to prepare students for the new and more synchronized expectations and outcomes for foreign language learning, teachers face a greater degree of accountability at different levels—school, district, state, and national—than has been the case in the past. Hancock stresses the “need for a close linkage between the ways in which instructors teach, test, and assess their students… …Initiatives are to be valued that assist teachers in reflecting about professional priorities and that help them make informed decisions” (1994, p. vii). With accountability so costly and so pervasive in American educational institutions today, it is critical that language teachers consciously focus on the reciprocal relationship that exists between teaching, testing, and assessment.

2.4 Performance Testing

Due to the nature of communicative language teaching, and the necessity to utilize authentic assessments in order to obtain a clearer understanding of students’ skills, teachers must now demand that students actually ‘perform’ the intended skills to be measured. While traditional testing relied mostly on objective multiple choice items which measured language knowledge rather than ability, (an approach which lent itself to
the use of statistics to achieve high reliability), purely objective items are no longer appropriate for assessing the communicative skills that students are expected to develop in today’s language programs (Shohamy, 2001). Testing is now performance-oriented and thus requires that students use the target language in order to communicate or interpret a “real-life” message. For example, students may be asked to read an authentic piece of writing in the target language, and then respond to what has been read. Or students may be asked to role play a situation from the target culture, using the target language to demonstrate their ability to communicate in places like a store, or at the airport, or to respond to a particular topic introduced by another speaker. With this approach, teachers must assess samples of student performance with a significantly different method for evaluating language ability, one that can be subjective in nature as opposed to the traditional objective measurement. Assessing student performances has resulted in a new set of controversies in the field of foreign language testing. These center around the following: how to develop appropriate and authentic tasks that discern between using the performance to assess the language, versus using the language to assess the performance; how to determine the criteria to be used for judging these assessments; and how to accurately determine the level of skills, and the degree to which we can predict a students’ future performance with the language (Bachman, 2002; Davies, 2001; Douglas, 2001; Elder, 2001; McNamara, 1997; Wiggins, 1994).

2.4.1 Performance Test Design and Evaluation

Wiggins (1994) argued for tests that help to improve performance as a result of their design and administration. According to him, authentic tests, or tasks, are those that include problems or questions worthy of importance, clearly outline the real-life task, are
multi-staged (as are our tasks in real world situations), offer transparent criteria and standards, provide for interaction between the candidate and the test administrator, and involve risk taking. He also stated that the judgment criteria should consider the sophistication of the answer and the candidates’ level of language, and should award the candidate for self-correction. Wiggins firmly believes that tests should no longer be shrouded in secrecy; rather, the candidates must know exactly what is expected, as well as the criteria for assessment, as ability is depicted as performing consistently when the task and criteria are known (1994, p. 69-85).

Bachman (2002) purports that the challenge of authentic assessment is great due to the complexity of its nature, and the difficulty faced when attempting to get at what we want to be able to predict from language learners. Whether the task itself is the assessment, or the task is a vehicle used to assess language, becomes an issue of concern for those developing and selecting tasks for tests. Attempts to isolate the language construct, develop the degree of difficulty for the task, and interpret learner interaction with the task, present additional causes for making it difficult to define the target language use domain, what exactly to do, and how the task will work. Similarly, Elder (2001) cited the same concerns with developing authentic tasks within the context of teacher tests in Australia. For example, Language for Specific Purposes tests (LSP) are those that intend to determine whether or not a candidate is equipped to perform a specific job or task. As with a performance test, learners must complete tasks in the target language; however, in LSP the tasks are specific to a particular job or duty that the learner will be carrying out in future employment. Like Bachman, Elder cites concerns for the ability to truly distinguish between the real world domain and test situation, and
the challenges faced in teasing out the components of ability or knowledge that the test measures. A pertinent question posed by Elder is whether or not the tasks can be separated out of the target language use domain for testing purposes. According to her, a common test technique like role playing can only depict the target language situation to a certain extent (2001, p. 149-170). Davies (2001) asks, how is it possible to create an unanticipated event from the real world? In order to try and create testing situations that depict real world tasks and provide information regarding a learner’s abilities, it is necessary to distinguish representative samples of the behavioral domain, while remaining keenly aware that these samples may be difficult if not impossible for the average test maker to demonstrate effectively. Given the challenges of trying to depict real world tasks in artificial testing situations, Davies believes authentic and performance tests should focus on a learner’s ability to manipulate language functions in a variety of ways, as this is what most often occurs in real world situations. For him this type of task would be most indicative of a learner’s skill level (2001, 133-147).

In the pursuit of tests that are indicative of future real-world performance, skill levels, and communicative ability, classroom teachers and test developers are confronted by the subjective nature of judging these performances. Clearly, this type of assessment is much less objective by nature, and therefore much more challenging to create fairly, and in turn judge by assigning a grade. Additionally, validity and reliability are compounded by the lack of objectivity that may accompany these types of tests and/or assessments.
2.4.2 Validity and Reliability in Foreign Language Testing

Testing in foreign language classes has long been linked to the concepts of validity and reliability. Validity is used to describe “the relationship between the evidence from test performance and the inferences about candidates’ capacity to perform in the criterion that are drawn from that evidence” (McNamara, 2000, p. 138). In other words, is the test taker’s performance on the particular skill, or behavior on the test, truly indicative of how s/he would perform in a target language situation outside the test that required use of that skill or behavior? Within the general domain of validity exist specific categories or types of validity:

- **construct** validity (the test measures the specific construct or definition of language ability being assessed)
- **content** validity (the content of the test is a representative example of the what the student has learned and needs to know)
- **criterion related** validity, which consists of **concurrent** validity (the test and criterion are administered at a relatively close point in time) and **predictive** validity (the degree to which the test can accurately predict the future performance of the candidate)
- **face** validity (the degree to which a test appears to measure what it is supposed to measure)
- **consequential** validity (the test produces results which are appropriate relative to the uses for which they are employed) (Hughes, 2003).

One of the concerns with the validity of performance tests is the degree to which a task realistically represents the relevant real world task. The test represents a series of
performances or a single performance that simulates or represents the relevant communicative behavior in the target language situation (McNamara, 1997, 132), but the accuracy of the task represented can vary. By the same token, it is impractical to assess a learner in the target language real world situation; therefore, it is necessary to manipulate the context of a test in order to recreate the task with as much authenticity as is permitted given the constraints with which we are faced. Due to the complexity of imitating real world tasks, the various other factors that play into a learner’s performance, and the subjective role that judgment plays in testing, validity can be a constant source of concern for testing in foreign language classrooms (McNamara, 2000).

The other key consideration, reliability, in general refers to the “consistency of measurement of individuals by a test…” (McNamara, 2000, p. 136). This represents the extent to which individuals will score the same or similarly on each administration of the test. Especially with performance testing, specific types of reliability include scorer reliability, referring to the degree to which a person evaluating a test scores accurately as compared to other scorers (inter-rater), and consistently scores accurately with respect to her/himself (intra-rater) (Hughes, 2003). Reliability may be influenced by the highly subjective nature of performance tests:

…performance assessment typically involves judgments of quality against some rating scale. This introduces new features of the assessment setting such as the raters themselves, who will vary in the standards which they use and the consistency of their application of those standards; and the rating procedures which they are required to implement (McNamara, 1999, p. 134).

In spite of the various factors of authentic tasks and performance tests that can affect the validity and reliability of tests, foreign language testing and assessment has forged ahead in the endeavor to assess learner’s skills and abilities in the target language. The
communicative assessment process is ambitious and complex, and requires an advanced understanding of not only potential inequities, but also ways to avoid these inequities when designing tests and assessments for the foreign language classroom (McNamara, 2000, p. 45).

2.5 The Influence of Tests on Instruction: Washback

Deliberate actions focusing on the communicative abilities that students are expected to develop, and that teachers are expected to facilitate in and out of the classroom, have generated the belief that the new means used to assess skills (authentic or performance based tests) can offer teachers more accurate opportunities to measure student abilities. These new means of assessment can include, but are not limited to, oral exams, letter writing, descriptions of pictures or story telling, debates, group projects such as newscasts, publishing a newspaper, or skits depicting different situations such as visits to the doctor’s office, the grocery store, the airport, etc. Typically these activities also reflect the national Standards for Foreign Language Learning. Because these assessments are more and more commonly built into the foreign language curriculum, they become part of daily class activities in an attempt to prepare students for the different types of tasks and activities that they will be asked to complete. As a result, tests and assessments of this nature are generally seen to render a positive influence on classroom instruction and practice (Valette, 1994; Wiggins, 1994), as what occurs in the classroom influences how assessment occurs, and vice versa. The connection between assessment and the methods and instructional techniques employed in the classroom is referred to as washback. Washback is “…the influence of language testing on teaching and learning” (Cheng, Watanabe, and Curtis, 2004, p. xiii). Messick elaborates:
… in the case of language testing, the assessment should include authentic and direct samples of the communicative behaviors of listening, speaking, reading, and writing of the language being learnt. Ideally, the move from learning exercises to test exercises should be seamless. As a consequence, for optimal positive washback there should be little if any difference between activities involved in learning the language and activities involved in preparing for the test. (Messick, 1996, p. 241-2)

Messick refers to “optimal positive washback,” which given the link between the activities that students perform in the classroom, and the task and performances that they are asked to perform for assessments, instruction and testing or assessment would appear to operate in complete synchronization. However, this is not always the case due to a variety of factors, the most notable in present educational settings being the importance given to national and state standardized exams, or district or school wide exams required to meet graduation requirements. The importance given to these exams can often influence teaching and instruction in negative ways. In other words, not all washback is positive. In particular, many language teaching specialists worry about the practice of “teaching to the test,” in which case sacrifices are made in what is taught or how it is taught in order to ensure that students can produce the type of language required on a test. This is referred to as negative washback. Negative washback has been observed and documented through such activities as teachers ceasing to use certain materials and activities that are not included on mandated tests, teachers including specific activities and or materials that are known to be on the tests, and giving mock tests to students using classroom time to administer these tests and discuss answers (Alderson & Hamp-Lyons, 1996; Shohamy, Donitsa-Schmidt, & Ferman, 1996).

It is undeniable to anyone who has ever experienced a test that instruction is usually influenced to some degree by a test. Questions such as “do we have to know this
for the test?”, “Is this going to be on the test?”, “How should I study for the test?”, “How will I know which answer you will want on the test?”, permeate all classrooms in every subject area. The age of the student does not matter; the bottom line is that everyone wants to be able to prepare for a test, and that means that it is necessary to know what is on a test. Because teachers have taken and given many tests, they are in a position to adapt their behavior and their teaching techniques to provide instruction and materials that will help to prepare students for the demands of the test. This is washback in its most basic form, often unconsciously appearing in daily class sessions.

2.6 Washback Research

The most commonly held view of washback is perhaps best expressed by J. Charles Alderson in his comments on his own research in the area of language washback as it began a decade ago:

I believe there is no longer any doubt that washback does indeed exist. …The question today is not ‘does washback exist?’ but much rather what does washback look like? What brings washback about? Why does washback exist? (2004, p. ix)

Investigating early washback, Wall and Alderson conducted a landmark study of washback in language testing in Sri Lanka from the middle of 1988 to the end of 1991. The purpose was to investigate the impact of a new ‘O-level’ English examination introduced in the Sri Lankan educational system in 1988. The importance of English in Sri Lanka is great, as many depend upon it to conduct international business, national business, and for social purposes. The ‘O-level’ examination, administered during what corresponds to North American 11th grade, is an exam that students must take upon culmination of their 11th year of study. A students’ grade on this examination determines
whether or not s/he will be allowed to continue with further study; or, if s/he enters into the workforce, if s/he will be suitable for sought-after jobs. In an effort to have students develop more practical English skills, the Ministry of Education developed new textbooks that placed a greater emphasis on reading and writing with a purpose, and on oral skills. In order to have these new innovations “taken seriously,” it was determined that a new ‘O-level’ test would accompany the textbook series to force the teachers and the students to meet the demands of the test, and consequently improve instruction and achievement (1996, p. 197).

The study included an evaluation of the O-level exam, focusing on validity and reliability of the test, and the impact that the test had on the classroom. The research was conducted in conjunction with Lancaster University (Wall and Alderson, 1996). In this study the researchers developed the basic framework for the Washback Hypotheses, which simply stated, posited that tests influence teaching. The specific hypotheses as developed by Alderson and Wall are as follows:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach.
4. A test will influence how teachers teach.
5. A test will influence what learners learn.
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching.
8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching.
10. A test will influence the degree and depth of learning.
11. A test will influence the attitudes to content, method, etc. of teaching/learning.
12. Tests that have important consequences will have washback.
13. Tests that do not have important consequences will have no washback.
14. Test will have washback on all learners and teachers.
15. Tests will have washback effects for some teachers and some learners, but not for others. (Alderson & Wall, 1993, p. 120-121 in Alderson & Hamp-Lyons, 1996, p. 281-2)
The study revealed that washback is a much more complex and challenging enigma than was first believed (Wall & Alderson, 1996). Furthermore, a simple, quantitative response to each hypothesis was impossible to articulate as a multitude of factors were found to contribute to washback in the classroom.

Attempts to document washback with empirical evidence have resulted in existing washback research that has been qualitative in nature. Qualitative research differs from experimental research or survey research in that the researcher is interested in “describing in detail all of what goes on in a particular activity or situation rather than on comparing the effects of a particular treatment…, or describing the attitudes or behaviors of people…” (Fraenkel & Wallen, 2003, p. 430). Additionally, because washback studies research how tests influence teaching and learning, it is necessary to examine tests that are a part of an existing curriculum:

Under artificial conditions, the test is likely to be perceived by the participants as having little educational consequence, which is unlikely in actual situations. These requirements necessitate using qualitative research methodology rather than a traditional experimental approach, although this does not preclude the use of quantitative data (Watanabe, 2004, p. 22).

Cheng and Curtis (2004) categorize the existing, limited research on washback into two major themes. These themes are: studies focusing on high stakes testing that tend to employ traditional (i.e., objective) types of tests in terms of design and format, and are generally assumed to exert negative washback; and studies focusing on a conscious change in a test or exam, modified in order to exert a beneficial influence on teaching and learning. Some studies may pertain to both categories, as they focus on
high stakes tests that were updated or modified in order to change the existing curriculum, and shift emphasis from or to different, yet specific language skills.

2.7 Washback Research in High Stakes Testing

High stakes tests are those that are considered to be tests or assessments used to influence or determine life-changing opportunities for individuals. In the academic context, tests of this nature are used specifically for such purposes as determining whether or not students will be allowed to study at the college level, permitted to graduate, awarded a scholarship, or even to be considered worthy of a particular distinction related to language skills (McNamara, 2000; Shohamy, 2001). Examples of these tests that have been the subject of studies include, but are not limited to, the Test of English as a Foreign Language (TOEFL), the Hong Kong Certificate of Examination in English (HKCEE), the International English Language Testing System in New Zealand (IELTS), and the National Matriculation English Test in China (NMET). These tests have been shown to have a distinct influence on the curriculum developed, the instruction provided in the classroom, the materials used to support instruction, and student behavior.

The TOEFL is a test used primarily in the selection of international students applying to study at universities in the United States. This test has served as the nucleus of further washback research, such as the study by Alderson and Hamp-Lyons (1996) that investigated the test and its washback effects on instruction. In this study, the researchers were not trying to "prove nor falsify any or all of the Washback Hypotheses, nor even to limit [the research to them]…. [rather the intention was] to simply try and explore and understand what was going on in TOEFL preparation classrooms" (Alderson & Hamp-Lyons, 1996, p. 282). The research project focused on teachers teaching TOEFL...
preparation classes, and used classroom observations, teacher interviews, and exploratory student interviews to gather data. The intention was to find out about the effect of the TOEFL examination on classroom instruction, teacher attitudes, and student attitudes, and compare the atmosphere of a TOEFL preparation class to a non-TOEFL class (Alderson & Hamp-Lyons, 1996). It was in this study that the Washback Hypotheses were further defined in an attempt to better understand the complexities associated with washback. The results of the study revealed what the majority of the research on high stakes testing continues to show: that the test (in this case, TOEFL) does affect teaching, and the non-TOEFL classroom is markedly different even with the same instructor. However, the process that results in these differences is much more complex than originally believed and involves many factors in addition to simply the test itself. These factors include not only characteristics of the test, but of the teacher and learner as well. This study revealed that washback is different for different teachers, in different settings.

Another well known high stakes test research project is The Washback Effect on Classroom Teaching of Changes in Public Examinations, also presented in greater detail as The Washback Effect of a Public Examination Change on Teachers’ Perceptions Toward Their Classroom Teaching, conducted by Liying Cheng (2000). This study focused on the changes introduced in 1993 to the Hong Kong Certificate of Education Examination in English (HKCEE). The changes introduced to the exam “reflected an integrated listening, reading, and writing exam, requiring students to perform simulated “real life” tasks, together with an increase in the weighting of the oral component” (2000, p. 92). Changes in the test resulted in changes in the role of the student as a language learner, demanding that s/he meet expectations that called for the development of skills,
and the completion of tasks. Although English teaching and learning in Hong Kong has been traditionally described “as centered on the 3Ts; test, teacher, and textbook” (Morris et al. 1996 cited in Cheng, 2000, p. 93), the purpose of invoking these changes in the test was to try and affect teaching and learning in an effort to improve it by moving toward more task based instruction. The purpose of the study was to determine the extent that the changes in the test actually resulted in classroom changes for teaching and learning English (Cheng, 2000; Cheng, 2004). The study revealed that “the change of the HKCEE toward an integrated and task-based approach showed teachers the possibility of something new, but it did not automatically enable teachers to teach something new” (Cheng, 2004, p. 164). In other words, just because a change is mandated via a test, does not guarantee that the intended changes and/or improvements will occur.

Further studies conducted by Qi (National Matriculation English Test, or NMET, in China, 2004), Watanabe (entrance examinations in Japanese universities, 2004), and Hayes and Read (International English Language Testing System, or IELTS, in New Zealand, 2004), also all focusing on high stakes testing, arrived at conclusions similar to those reported in the previously mentioned studies. Results indicated that different teachers exhibit varying degrees of washback, teacher differences appear to be as important a factor as testing factors in relation to washback, and implementing a test or examination does not guarantee that the desired changes and/or effects will occur. In summary, these findings suggest that washback does occur and directly influences teaching; however, as put succinctly by Hayes and Read in their study, “[they] did not seem to be the kind of positive effects envisaged at the outset of [the] study” (p. 111).
Thus, while washback does seem to exist, there is continued uncertainty as to the degree to which it is positive, and how to achieve positive washback.

2.8 Washback Research on Tests to Drive Innovation and Change

The second theme of the research on washback considers the use of tests to drive curricular change. This idea is not new; in fact, Davies (1985) previously suggested using tests as a means to provide impetus for curricular change and teacher change, stating that a test was the fastest and most sensitive method, as it is the most controllable, and hard to ignore means to bring about a change. Tests are an integral part of education. Davies (1985, cited in Wall, 1997) cautioned, however, that change should not be so abrupt or rapid that the teacher and the syllabus would be unable to catch up with it.

Washback studies focusing on this theme have examined tests that have been implemented via language policies, ministries of education, and/or university systems as a means to force teachers to provide increased instruction in focused skill areas. This is consistent with an observation made by Shohamy, Donitsa-Schmidt, and Ferman (1996): “The power and authority of tests enable policy makers to use them as effective tools for controlling educational systems and prescribing the behavior of those who are affected by their results – administrators, teachers, and students” (p. 299). A variety of language tests in foreign settings, and standards based tests in the United States, have been produced in an attempt to control and change the curriculum. Shohamy (1996, 2001) has done extensive research in Israel on the effects of such tests on education and instruction. The two tests that were the focus of her research for the 1996 study were the Arabic Second Language Test (ASL) and the English Foreign Language Test (EFL). The ASL test was designed for seventh, eighth, and ninth grade students studying Arabic as a foreign
language, and the EFL test was an oral test administered to all twelfth grade students graduating from high school. The ASL test was introduced ..... “to raise the prestige of the Arabic language, to equalize levels of teaching in schools, to force teachers to increase the rate of teaching, …and to increase the motivation of both teachers and students to teach and learn Arabic” (Shohamy, Donitsa-Schmidt, & Ferman, 1996, p. 301). The EFL test was introduced ..... “to increase the emphasis on teaching oral language in the classroom in order to upgrade students’ oral proficiency” (Shohamy, Donitsa-Schmidt, & Ferman, 1996, p. 301). These tests had a significant impact throughout the Israeli education system on the teaching and learning in classrooms. Results of Shohamy et al’s research indicate that in many regards the tests drove the curriculum and classroom instruction, and not always in educationally sound ways. For instance, they observed such teacher and student behaviors as reverting to extensive review of test-related material and avoidance of new material, using worksheets that were a direct reflection of previous tests, focusing predominantly on test-like activities in the classroom, and exhibiting a high level of tension as well as a high degree of motivation to master the test-related material on the part of both the teachers and the students. There was considerable negative washback observed arising from the introduction of these standardized tests (Shohamy, Donitsa-Schmidt, & Ferman, 1996).

Another test introduced to drive innovation and change was the high stakes Hong Kong Certificate of Education Examination in English (HKCEE), the focus of Cheng’s study (2001, 2004). Though it resulted in the introduction of a task-based approach to testing through the integration of listening, reading, and writing skills, results indicated that washback occurred in the teaching and learning, but had little effect on classroom
methodology. Cheng concluded that “…ways in which teachers perceive teaching and learning might change, yet they are not likely to influence the ways in which they teach. The change to a new exam has informed the teachers about what they might do differently, but it has not shown them how to do it” (2002, p. 108).

The results of these studies focusing on the use of tests to impart change, indicate that change is much more complex and sophisticated, and for it to occur, more is required than simply integrating a test into a curriculum. Concrete evidence supporting this idea can be found in present day classrooms throughout the United States, where state mandated testing in the core subject areas (e.g., math, science, writing) as a result of government initiatives is being used to drive curricular change in public school institutions. The washback effects of these tests are both positive and negative, and need to be further researched, publicized, and scrutinized in order to determine whether the tests are serving the purpose they were created and implemented for, and to ensure that teacher training and preparation articulates with the classroom demands commanded by these tests.

2.9 **A Future for Washback Research in the United States**

The most significant washback studies have been conducted in various countries outside of the United States in high stakes testing environments, as these are places where one finds the most common use of large scale language tests. Although there are currently no federally mandated foreign language tests in the United States, there are some states that do participate in standardized language testing on some level. One previously mentioned example is New York, which instituted the New York Regents
Exams. These examinations are offered twice a year (January/June) in French, German, Hebrew, Italian, Latin, and Spanish ([www.nysedregents.org/testing/hsregents.html](http://www.nysedregents.org/testing/hsregents.html), retrieved March 10, 2004). These tests evaluate reading, writing, speaking, and oral skills.

In this new test-oriented environment, teachers are required to work with new materials and guidelines in order to help students achieve pre-determined benchmarks of ability or performance that will permit them to progress through different grade levels and be eligible for graduation from high school. Although foreign language classes are not subject to mandatory examinations, many programs have implemented specific criteria that students must meet. For some programs these criteria must be met in order to progress to the next level of language study, and they may also include language requirements set for graduation. The criteria developed for many programs detail skills and abilities that students must demonstrate at a particular level, or upon successful completion of a course sequence. Implementing these benchmarks within a program may be viewed as the preliminary stages of high stakes accountability, additionally serving to spark curriculum innovation and change within the classroom. Undoubtedly the washback effects of these demands will be observable, but the extent of the positive and negative effects of this washback is yet to be determined.

The principal focus of the previously cited studies, as well as this study, is washback in foreign language classes. However, in an era when high stakes testing is gradually becoming the law of the land in the United States, as dictated by the federal *No Child Left Behind* legislation, the research reported here has important implications for a wide range of disciplines beyond foreign languages because of the prevalence of testing. As a result
of recent and ongoing education reform in the United States, it appears that policy makers will continue to view assessment, particularly in the form of high stakes, standardized tests, as an appropriate method for spurring change in curricula and instruction. According to Shohamy, the appropriateness of this approach to school reform is debatable:

Tests…can open or close doors, provide or take away opportunities, and in general shape the lives of individuals in many different areas. …tests are used as a method of imposing certain behaviors on those who are subject to them. Tests are capable of dictating to test takers what they need to know, and what they will learn and what they will be taught. The use of tests as disciplinary tools means that test takers are forced to change their behavior to suit the demands of the test. …Using the tests as disciplinary tools is an extension of the manipulation of tests by those in authority- policy makers, principles, and teachers- into effective instruments for policy making. It is the realization that test takers will change their behavior in order to succeed on tests that leads those in authority to use tests to cause a change in behavior in accordance with certain priorities. (2001, p. 16-17)

Students in foreign language classes are becoming increasingly more diverse and are studying languages for a variety of reasons. The undergraduate, intermediate level classroom that is the focus of this study provides a particular challenge for both students and educators, as this is the level where students are first expected to use the target language and refine their skills (Rava, 2000). The transition to this level can be problematic for both students and teachers, especially when expectations and educational objectives for each group are not always clearly articulated. As Rava (2000) explains, “faced with the challenge of striking a balance between institutional language requirements and preparation for more advanced language and literature studies, faculty and students alike …often view the intermediate curriculum as the ugly stepsister” (p. 342). If foreign language education aims to meet the needs of students through language
programs, and to prepare them for success in an increasingly global society, then
educators must seek to discover ways by which teaching and assessment can be linked as
seamlessly as Messick describes. To do so, educators and policy makers must continue
to increase the knowledge of washback in the classroom. Tests that are a regular part of
the curriculum are considered low stakes tests, and virtually a part of all classes in the
majority of academic institutions in the United States. Consequently they provide an
opportunity to study washback as it occurs in a variety of educational settings. As seen,
research to date, especially in the United States, has failed to provide a clear and
convincing picture of the extent to which and ways in which washback is manifested in
foreign language instruction. By focusing on a low stakes testing environment, this study
of intermediate level students of Spanish as they move from one course to another where
there is a unified testing system in place, should help to bridge this gap, especially since
the tests and assessments used in the modern foreign language classroom focus more on
the Standards, and try to include a range of student skills and abilities. A focus on skills
and abilities is also something that will be an important component of any national or
state standardized test that is developed and implemented for foreign languages.
Additionally, this study can provide insight for other disciplines with respect to the
influence that tests have on teachers’ instructional behaviors and corresponding student
behaviors.
CHAPTER 3

METHODS

3.1 Introduction

In the simplest of terms, washback is the influence of testing on teaching and learning. As Cheng (2004) states, “in spite of growing literature on washback, there is relatively little empirical research in this area. …In order to further our understanding, we need to look at the phenomenon in a specific education setting by investigating in depth different aspects of teaching and learning” (p. 148). The purpose of this study was to investigate and describe washback in a low stakes testing situation in the intermediate Spanish foreign language classroom.

Teachers from the Ohio State University Department of Spanish and Portuguese teaching Spanish 104 for the first time served as the primary subjects in this initiative. Through qualitative and quantitative methods, that included a total of sixteen class observations, and four interviews per teacher, a teacher questionnaire, and student questionnaires, the two part study provides the details of washback observed in the classroom. The study describes washback as it related to observable, changed teacher behaviors that directly reflected the Spanish 104 test. It also examines how these changed teacher behaviors may have influenced student behaviors.
The instruments used for gathering data focused on eliciting information related to the impact of increased knowledge of the Spanish 104 testing program on a teacher’s instructional techniques and strategies. The instruments also provided additional information about past teaching experience, teaching style, previous language learning experience, and teaching beliefs as they related to instructional behaviors. The detailed description of washback informs the foreign language profession as it seeks to better understand how tests influence teaching and learning. With an increased understanding of the influence of tests on instruction, testing practices, instruction, and teacher training and professional development programs, may be improved.

3.2 Design and Procedures

3.2.1 Research site

The research site was the Department of Spanish and Portuguese at the Ohio State University, with data gathered from purposefully selected sections of Spanish 104, an intermediate language course.

3.2.2 Participants

The primary participants for the study were five teachers teaching in the intermediate language program (Spanish 104) for the first time in the Autumn 2004 quarter. The participants were selected from a population of Graduate Teaching Assistants and Lecturers in the Department of Spanish and Portuguese at The Ohio State University. Teachers were informed of the purpose of the research and asked for a commitment to participate in the study, which required them to teach Spanish 104 for the Autumn and subsequent Winter quarter.
The teachers selected all had at least one year of teaching experience in the The Ohio State undergraduate Spanish elementary language program. Although preference was for teachers with a minimum of two years prior teaching experience, the extent to which this was possible depended upon the teaching schedule. Because it was necessary to have teachers who would be teaching Spanish 104 for the first time in the Autumn quarter, and who also agreed to teach that same course in the Winter quarter, this condition was not met in two cases. The expectation was that teachers who had prior experience teaching elementary level courses, and with the testing program at that level, would have limited knowledge of the unique aspects of the intermediate level course and its corresponding testing program.

The secondary participants for the study were the students from the classes of the teachers who participated in the study. Although not a main focus of the research, the choice to include student data was informed by previous research and current practice. In previous washback research, the majority of the studies have not included student data in the research. Ferman’s (2004) study of the EFL test implemented by the Israeli Ministry of Education appears to be one of the only studies that included student data. He found that students’ washback behaviors appeared to be influenced by the teachers’ instructional behaviors with respect to the test. This seemed an important aspect to consider for the Spanish 104 study, as the intent was to observe how teacher behaviors changed over a period of time. Additionally, current educational practice uses student performance as a common way to judge teacher efficacy. If teachers can be judged on how well their students perform on tests, it is relevant to gather data to determine whether
or not teacher behaviors related to the tests, in turn influenced student behaviors with respect to the tests.

3.2.3 The Researcher

As a participant observer, the researcher participated in the activities of the group being studied, while also making it clear that she was conducting research. Participation included engaging in activities with student groups, and whole class activities, and answering questions or providing information when called on by the teacher. In this study, the participant observation was overt, “as the researcher [was] easily identified and the subjects [knew] that they were being observed” (Fraenkel & Wallen, 2003, p. 451).

The researcher was an experienced teacher with twelve years of Spanish teaching experience in public schools, and three years of Spanish teaching experience in higher education. The researcher was also an experienced Spanish 104 teacher in the Department of Spanish and Portuguese, and she was teaching a section of Spanish 104 during the Autumn quarter of the research project.

The researcher’s experience with foreign language testing included developing district wide foreign language placement tests, developing school based foreign language tests, scoring state wide Foreign and Bilingual tests required for teacher certification, and scoring Advanced Placement Spanish Language tests for the Educational Testing Service®.

In the Department of Spanish and Portuguese at Ohio State, the researcher had previously worked scoring the Spanish Listening and Reading Proficiency Exam, she coordinated the elementary level testing program, and worked with the Spanish 104 Program Coordinator on the tests for Spanish 104. Although it was not the researcher’s
intention at the onset of the study to be a participant observer in the research, it was unavoidable. Because of the researcher’s extensive experience with teaching and testing, and her experience in the Spanish 104 program, it was impossible not to become a participant in the research. Both teachers and students drew the researcher into the project.

3.3 Intermediate Spanish 104

3.3.1 Course Description

Spanish 104, as described on the website for the Department of Spanish and Portuguese (www.sppo.ohio-state.edu), is an intermediate language class that places a strong emphasis on reading and writing through the use of authentic literature and reading samples, while continuing to build students’ language skills for listening, reading, writing, and speaking. Spanish 104 provides a challenge for the students who have passed through the Spanish 103 program, where there is less focus on reading and writing. Although there are numerous examples of authentic materials at the 103 elementary level that focus on the development of the skills (listening, reading, writing, and speaking), the volume of the content, along with the predetermined course calendar, the varied teaching repertoires of the instructors, and the standardized department tests that contain significant structural content, can limit the amount of instruction devoted to skill development, especially in the areas of reading and writing.

The Spanish 104 calendar (i.e., the syllabus) for the entire quarter is published prior to the start of each quarter. It includes the content to be covered in each class session, as well as suggested homework and the schedule for all required assessments. The calendar content is based on the curriculum and the testing schedule.
3.3.2 The Spanish 104 Testing Program

In Spanish 104, there are three standardized department chapter tests, and a final exam during each quarter. The tests and the final exam account for fifty five percent of the student’s final grade for the course. Test 1 and Test 2 focus on material from two chapters in the text *De paseo* (Long and Macián, 2003), and Test 3 focuses on one chapter. The final exam, although cumulative, focuses predominantly on the last chapter studied in class, with a structure section that includes grammar concepts from other chapters. For every test, there is a review page that provides information about the material from the text to be included on that exam. All students and teachers have access to this page on the Spanish 104 website.

Each chapter test includes a listening comprehension section, a reading comprehension section, a textbook reading section, and a language/structure section. For listening comprehension, students listen to a passage in Spanish and write answers to the questions in English. For reading comprehension, students read a never before seen passage in Spanish and write answers to the questions in English. For the textbook reading section, students write answers to the questions about authentic readings from the textbook studied in class. Questions are all in Spanish and the students are required to answer in Spanish as well. Students must provide details, descriptions, summaries, and general information to demonstrate comprehension of the readings. Instructions on the test state that students are penalized for writing words in English. For the language/structure section, students provide grammatically accurate forms of structures from the recently studied chapters.
The sections on listening and reading comprehension, and on language/structure, are similar between the elementary and intermediate programs. The contrast between the Spanish 103 (elementary) and the Spanish 104 (intermediate) testing program is observed most emphatically in the section on the textbook readings that are a significant part of the course curriculum. A breakdown of the distinct sections, with the point distribution included on a typical Spanish 103 test (the final course of the elementary program), and for a typical Spanish 104 test (the first course of the intermediate program), is provided in Table 3.1.

<table>
<thead>
<tr>
<th>Spanish 103</th>
<th>Points</th>
<th>Spanish 104</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comprehension</td>
<td>20</td>
<td>Listening comprehension</td>
<td>17</td>
</tr>
<tr>
<td>Dictation</td>
<td>10</td>
<td>Reading comprehension</td>
<td>18</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>20</td>
<td>Textbook reading questions</td>
<td>33</td>
</tr>
<tr>
<td>Structure and grammar</td>
<td>35</td>
<td>Language and grammar</td>
<td>32</td>
</tr>
<tr>
<td>Structure embedded in writing</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Culture</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL POINTS</strong></td>
<td>100</td>
<td><strong>TOTAL POINTS</strong></td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.1. Spanish 103 and Spanish 104 test sections and point distribution

As illustrated in the table, the testing program at the Spanish 103 level demands that students demonstrate comprehension skills for forty percent of the test, while at the Spanish 104 level, students must demonstrate comprehension skills for sixty eight percent of the test. Additionally, on the Spanish 104 test, thirty three percent of the test demands that students read and write in Spanish to demonstrate their comprehension skills (“textbook reading questions”). It is precisely this focus on skills (predominantly in the
area of reading and writing) that contributes to the conditions resulting in washback behaviors that this study investigated.

The chapter tests that constituted the major focus of this study are considered low stakes tests, which means that the consequences of passing or failing the tests are not “perceived as connected and as responsible for a number of additional events, contexts, and consequences” (Shohamy, 2001, p. 8). Regardless of the stakes associated with the Spanish 104 chapter tests, they are still important to students’ success in Spanish 104. The tests contribute 55% to the student’s final average (25% for the three chapter tests, and 30% for the final exam). It is possible for a student to fail one or two of the chapter tests and still pass the class provided s/he does well on the other assessments. However, it is likely that if a student fails all of the chapter tests, s/he may not pass the course.

The students and the teachers demonstrate behaviors that indicate the importance that they give to these low stakes tests. For example:

• Teachers spend class time making sure that students understand the format of the tests and that they know the specific material that will be included on the tests.
• Teachers make sure that students understand how to access the review pages for the tests, and make themselves available for extra help.
• Students ask frequent questions, and demonstrate anxiety with respect with upcoming chapter tests.
• Students ask their teachers for extra help before a test.

Although the reasons behind these behaviors may differ, students want to be successful on the tests, and teachers want their students to be successful on these tests.
Another aspect that may contribute to the importance given to tests by teachers and students is that in the Department of Spanish and Portuguese, a grade of 64 or below is considered failing, whereas in other educational settings, the distinction between passing and failing is a grade below 60. Although the students who have progressed through the Spanish language program at The Ohio State University are aware of the grading policy, for the students who test directly into Spanish 104, this is distinct. This further plays a part in the test anxiety that both students and teachers demonstrated in this study, as observed in class sessions, and as reported in questionnaires from both teachers and students.

Additionally, in contrast to other washback research studies, the tests in this study were not implemented to change instruction, or as a result of educational policy changes. Rather, these tests were a part of the existing Spanish 104 curriculum.

3.4 Instrumentation

The methodology for this study was qualitative in nature, employing a variety of techniques and methods to gather data on washback. This means that there is no “treatment”…, nor is there any “manipulation” of subjects. The participants in a qualitative study are not divided into groups, with one group being exposed to a treatment or some sort and the effects of this treatment then measured in some way. Data are not collected at the “end” of the study. Rather, the collection of data in a qualitative research study is ongoing. The researcher is continually observing people, events, and occurrences, often supplementing his or her observations with in-depth interviews of selected participants and the examination of various documents and records relevant to the phenomenon of interest. (Fraenkel & Wallen, 2003, p. 435)

Data analysis in qualitative research is characterized by description. “Essentially [it] involves analyzing and synthesizing the information the researcher obtains from various
sources (e.g., observations, interviews, documents) into a coherent description of [what has been] observed…” (Fraenkel & Wallen, 2003, p. 435).

3.4.1 Observations

In qualitative research, naturalistic observations offer the researcher a means to observe individuals in their natural settings. “The researcher makes no effort whatsoever to manipulate the variables or to control the activities of individuals, but simply observes and records what happens as things naturally occur” (Fraenkel & Wallen, 2003, p. 451).

For this study, the researcher recorded classroom observations by using a method adapted from Jon Saphier’s model from The Skillful Teacher for administrators. This method was chosen as the framework for the development of the Ohio State University observation note taking sheet because it allowed the greatest space and detail for commentary. It also offered an effective way to accurately record the maximum amount of information describing what occurred in the classroom, and had been used successfully by the researcher in a multitude of classroom observations. A copy of the form is found in Appendix A.

The format of the observations sheet allowed the researcher to record everything said by the teacher and the students, with dedicated columns for each. The researcher also recorded the time that each event occurred in the classroom, which enabled her to calculate the percentage of class time spent on each activity, and then calculate how much time teachers spent on specific topics overall.

Reliability of the Spanish 104 study’s observations was checked against observations recorded by an independent researcher. This independent researcher was an experienced teacher, who speaks Spanish fluently. She is currently the Supervisor for
students in the Master’s of Education Program in Foreign and Second Language Education for the College of Education at the Ohio State University. The independent researcher and the Spanish 104 study researcher simultaneously observed one of the study teachers teaching her Spanish 104 class and recorded the observation on the observation note taking sheet used in this study. The Spanish 104 researcher’s observation and the independent researcher’s observations were compared against each other and revealed 100% agreement. No observational discrepancies were revealed.

3.4.2 Interviews

In qualitative research, “interviewing (i.e., the careful asking of relevant questions) is an important way for a researcher to check the accuracy of—to verify or refute- the impressions he or she has gained through observations” (Fraenkel & Wallen, 2003, p. 455). For the purpose of this study, as in other washback research, interviews allowed access to reasons behind some of the behaviors observed in the classroom during the research (Watanabe, 2004, p. 23). The format of the first interview followed a modified version of the interview form used by Qi (2004) for her study of the washback effects of the National Matriculation English Test in China. This interview protocol was chosen as the model because of its construction. It elicits the desired background information about teachers, and it is of appropriate length for this study, demanding only 15-30 minutes of the participant’s time. The questions are direct, asking for pertinent information such as the teacher’s plans for the course, and her perception of the anticipated test. Initial modifications were made in order to tailor the questions to reflect the nature of the Spanish 104 course at the Ohio State University, the teachers, and the testing program. The interview questions were reviewed by a panel of experts that
included, but was not limited to, past 104 course teachers, the Academic Program Specialist and 104 Supervisor, and the Nighttime teaching supervisor. Questions were then field tested prior to use and necessary changes were implemented.

The second interview modeled the first interview; however, it operated more like a semi-structured interview (as opposed to a tightly structured one) in an attempt to allow “interviewees maximum freedom, [so that] ample and perhaps unexpected information [might] emerge” (Ferman, p. 195). In other words, the researcher had a set of questions pertaining to how the testing program may be affecting teaching, but in order to allow the participant freedom of expression, and to avoid “leading” the participant with focused questions, the researcher engaged the participant in a dialogue, instead of a question and answer session. The second set of interview questions was developed by the researcher, and then reviewed by the same panel of experts that included but was not limited to, past 104 course teachers, the Academic Program Specialist and 104 Supervisor, and the Nighttime teaching Supervisor. This set of questions was also field tested prior to use and necessary changes were made.

Working from the framework of the first two interview instruments and the questions from the student and teacher questionnaires, the researcher developed the third set of interview questions and also questions for the second quarter interview. For the final interview of the first quarter (the third interview), the questions were designed to elicit information about changes that had occurred in the teacher’s behaviors as a result of the increased knowledge of the testing program. There were also opportunities to gain specific information that could offer insight as to why these changes might have occurred. Teachers were asked to respond to their perceptions and beliefs about the tests
and their classes. As a result of their first quarter experience, the teachers had a different perspective. For that reason, although questions were similar in nature to questions that the teachers had already been asked, the responses were different because of their personal experience. Teachers were allowed the freedom to express their personal reactions to their first quarter of teaching the course, and to any instructional changes that they made either during the Autumn quarter, or that they anticipated making in the upcoming Winter quarter. These behavioral changes included, but were not limited to:

- Introducing ancillary materials different than those offered in the text.
- Eliminating or reducing the use of ancillary materials not specifically offered in the text.
- Increasing focused class time on reading activities.
- Increasing focused class time on writing activities.
- Reducing class time on speaking activities.
- Increasing class time on focused grammar instruction, and/or decreasing class time on focused grammar instruction.

The final study interview questions were to gather information about the teacher’s perspective on changes in instruction between the Autumn quarter and the Winter quarter. The purpose was to see whether the teacher perceived these changes in the same way that the researcher did, and whether the teacher acknowledged the change in her behaviors as a result of washback from the testing program. In this interview as the others, the teachers were allowed the freedom to elaborate.
3.4.3 Teacher Questionnaires

After teachers administered and scored Test 2 in the Autumn quarter, they were given a written questionnaire to complete. The questionnaire was designed to elicit their response to the testing program and their perception of the extent to which their teaching adequately reflected or matched the expectations of the test. The questionnaire also asked about teachers’ beliefs with respect to language learning. This was to gain further information that might help to explain some of the teachers’ instructional decisions. The teachers responded to the questions by rating prompts on a Likert scale allowing for quantitative data in the form of descriptive statistics.

The questionnaire model was an adaptation of Cheng’s (2004) teacher questionnaire used in her washback study of the Hong Kong Certificate examinations in English. This model was chosen because it had been successfully used in a previous washback study where teachers were expected to implement instruction reflecting new testing procedures. Although Cheng’s was a high stakes context and Spanish 104 is a low stakes context, the premise of new testing procedures was similar. The questions from Cheng’s model were modified to reflect the needs of The Ohio State study.

3.4.4 Student Questionnaires

After each test, and upon completion of the Spanish 104 final exam, the students from each teacher in the study completed a questionnaire. The brief questionnaire after each test solicited information about student study habits. The purpose was to see if students appeared to change their study habits as a consequence of teachers changing instructional habits and instructional emphasis.
The final exam questionnaire that students completed was a modified version of Ferman’s (2004) student questionnaire as used in the washback study of a national EFL oral matriculation test introduced by the Israeli Ministry of Education. Again, although Ferman’s model was designed specifically for students involved in a high stakes test washback study, it provided a format that elicited the relevant information from the students regarding testing, in a timely fashion. The questionnaire asked students about their perceptions of their teacher, and whether they believed that her behaviors reflected testing demands. Students were also asked questions about their perceptions of the test.

Ferman’s questionnaire was modified to match the needs of The Ohio State University study. It asked for basic personal information, such as year in school, major, age, Spanish courses taken in The Ohio State University Spanish language program, but not names. Students were also asked to grant permission to access test scores from Spanish 103 and Spanish 104 for future washback research endeavors.

3.4.5 Pilot and Field Testing

The teacher interview questions and questionnaires were pilot tested prior to the start of the Autumn 2004 quarter to provide information regarding the validity and reliability of the instruments for the purposes of the research. A panel of experts that included, but was not limited to, the Spanish Language Program director, experienced Lecturers who had previously taught Spanish 104, the Portuguese Language Program director and Coordinator of Spanish 104, and current Academic Advisors who were also former Spanish 104 instructors, evaluated the instrumentation in the pilot test. This process led to advancement for the field test. The field tests were conducted with teachers who taught Spanish 104 for the first time during the Summer quarter of 2004.
Student questionnaires were pilot tested in the same manner, and field tested with the researchers’ Spanish 104 students who were not part of the study. As a result of the pilot and field tests, further modifications were made in order to clarify questions, increase the precision of the questions, and make any other changes deemed necessary.

3.5 Procedures: Autumn Quarter

The following section describes the procedures that were used in the first part of the study from September to December. The data gathering included the following:

- **Teacher focused**: 8 observations, three interviews, and a questionnaire.
- **Student focused**: 3 questionnaires, one following each chapter test, and a final exam questionnaire.

3.5.1 Observations

Upon selection of teachers for the study, specific class sessions were chosen for observation. The classes were selected based on the lesson scheduled for that day, and its relationship to the material included on the impending test. This offered a focused opportunity to investigate potential washback behaviors. Because the researcher had direct knowledge of the course, access to the tests, and worked with the test writer for the examinations, she was able to identify the lessons and materials that directly impacted the test. Certain components of the curriculum are covered explicitly on the tests, and others are not included. The observations were of lessons that focused on material specifically included on the tests, and material that the researcher knew would not be on the test, or at least not directly tested.

An observation schedule was devised and distributed to the teachers during the first week of the Autumn quarter. Observations consisted of ‘test impacting’ and ‘non
test impacting’ observations. ‘Test impacting’ observations were lessons that included test content, and ‘non test impacting’ were lessons that included material or activities that were not part of test content (vocabulary, speaking activities, role playing activities, culture, etc.). ‘Non test impacting’ observations were included to try and minimize the influence on a teacher’s behavior who assumed that the observations were tied solely to material directly impacting the test. ‘Non-test impacting’ observations lent balance to the observation process and minimized the possibility of the observations contaminating the data gathering process. This also gave the researcher opportunities to observe a broader range of the teacher’s instructional behaviors. The researcher designated all observations as ‘test impacting’ or ‘non test impacting’ in her notes, and in the section on Classroom Observations in Appendix D.

Observations were conducted in the following manner. During the first observation for each class, the researcher was introduced, and the purpose of the researcher’s observations was explained to all students. Students were encouraged to approach the researcher with any questions or concerns about the study at any time. For the subsequent pre-scheduled observations, the researcher observed each teacher participant teach the same lesson as described on the calendar to her class. The researcher sat in the back of the class, or to the side of the class so as not to be a disruption to the class, and to limit contact and interaction with the students. Although it was never the researcher’s intention, after the first couple of observations, students began to interact with the researcher on a regular basis about classroom issues, curriculum topics, and for general questions about the target language. At times the teachers called
on the researcher during class as if she were a regular member. Both teachers and
students came to see the researcher as a class participant.

The researcher recorded information such as the teacher behaviors exhibited
during the class, the student behaviors exhibited as a result of instruction, the
content/material, and the time allotted to different activities. Comments with regard to
emphasis placed on content/material directly related to the test, and any audio-visual
sources utilized in the lesson were also noted. The class observations also recorded
descriptive information about the class such as the time of day, class interaction prior to
class commencement and/or post class culmination, and any behaviors or occurrences
that the researcher felt might be worthy of potential analysis because they appeared to be
behaviors indicating washback. The information from the classroom observations, such
as the division of instructional time for each lesson for each teacher, and which lessons
were categorized as ‘test impacting’ and ‘non test impacting’, are in Appendix D.

3.5.2 Interviews

Along with the observations of ‘test impacting’ and ‘non test impacting’ lessons,
teachers participated in three interviews during the Autumn quarter. The first interview,
conducted at the teacher’s convenience, occurred prior to Test 1. This interview focused
on the teacher’s plan for the course, and what her perceptions were of the differences
between Spanish 104 and other courses that she had previously taught. The second
teacher interview occurred after Test 1 was administered, but prior to Test 3. This
interview focused on how the teacher perceived that the tests in Spanish 104 changed her
instruction, if at all, since the beginning of the quarter. A final Autumn quarter interview
was held after teachers had administered and scored the Spanish 104 final exam. This
final interview elicited responses about behavioral changes affecting instruction that teachers made as a result of their increased knowledge of the test and the testing procedures for Spanish 104.

All teacher interviews were recorded on audio tape and then transcribed by the researcher. The interview transcriptions were coded and analyzed by the researcher according to teacher beliefs that influenced behavior patterns observed in the classroom. Teacher beliefs that appeared to influence behaviors were coded according to those that were observed most frequently, thereby dominating instruction; those that reflected washback behaviors from the testing program; and those that were discussed in interviews but were not observed frequently. Copies of all interview questions are in Appendix B.

3.5.3 Teacher Questionnaire

Following the administration and scoring of Test 2, the teachers completed a written questionnaire. The purpose of the questionnaire was to gather data about the teacher’s perceptions of the testing program in Spanish 104, the transition between the elementary Spanish language program and the Intermediate Spanish language program, and teacher beliefs about language learning. The questionnaire asked teachers to respond to questions by rating prompts on a Likert scale. This allowed for quantitative analysis of the teacher perceptions and beliefs with regard to Spanish 104 with mean responses, and highest rated responses. The teacher questionnaire is in Appendix B.

3.5.4 Student Questionnaires

Students were asked to complete a short questionnaire after each test. The questionnaire asked students to indicate the amount of time spent studying for each
section of that test, and their overall study time. The intention was to calculate the student study times and then compare them to teacher’s instructional behaviors.

Upon completion of the Spanish 104 final exam, each instructor encouraged students to complete a longer, more detailed questionnaire. No student was forced to complete the questionnaire. The purpose of this questionnaire was to gather information about students’ perceptions of their teacher’s instructional behaviors as they reflected expectations for the test, and students’ reactions to their increased knowledge of the testing program as the quarter progressed.

The questionnaire also asked students’ permission to access their test grades only from Spanish 103, and their test grades only from Spanish 104. This was done for future washback research to calculate a series of anonymous test means, including student test score means and class test score means, for both Spanish 103 and Spanish 104.

In summary, Autumn quarter research included eight classroom observations of each teacher, three teacher interviews, one teacher questionnaire, three short student questionnaires, and one comprehensive student questionnaire.  

3.6 Procedures: Winter Quarter

The following section describes the procedures that were used in the second part of the study from January to March. The data gathering included the following:

- **Teacher focused:** 8 observations, one interview.
- **Student focused:** 3 questionnaires, one following each chapter test, and a final exam questionnaire.
3.6.1 Observations

During the Winter quarter, the same teachers from Autumn quarter were observed teaching the same lessons. The same instrument for classroom observations was used, and the data was recorded in the same way previously described. The purpose of the observations was to provide the researcher with information about changes in teacher behaviors as they occurred between Autumn quarter and Winter quarter classes.

3.6.2 Teacher Interview

The second quarter teacher interview was an adaptation of the third teacher interview used during the Autumn quarter. Because the teachers were now familiar with the testing program, the only interview conducted during the Winter quarter was either after the administration and scoring of Test 3, or before or after the Spanish 104 final exam. This time was chosen for the interview because the teachers had completed the testing cycle for the course, and the content for the tests was identical, even though the specific format usually varied slightly. This also gave the researcher the opportunity to observe all of the same lessons observed in the Autumn quarter a second time, which therefore enabled her to have more information about the individual teacher’s behaviors before the interview.

This structured interview was intended to solicit data regarding the teacher’s perception of how her teaching changed in the Winter quarter as compared to the Autumn quarter as a result of the knowledge of the testing program. Questions were designed to uncover the reasons that teachers used certain activities in class, the information that they communicated to the students about the test and how to prepare for it, and how they decided to emphasize certain materials in the course more than other materials.
The final teacher interview was also recorded on audio tape and then transcribed by the researcher. The researcher coded and analyzed the interview transcriptions in the same manner as described for the Autumn quarter interviews. The final interview questions are in Appendix B.

3.6.3 Student Questionnaires

Students in the Winter quarter followed the exact same procedure for the short questionnaire. After each test, the students completed a questionnaire indicating the amount of time that they spent studying for each section of that test, and their overall study time.

The students from the Winter quarter classes of the study teachers were given the same final questionnaire that the students in the Autumn quarter completed after the Spanish 104 final exam. The results of the questionnaires from the Autumn quarter were compared with the results of the Winter quarter to determine similarities and differences, and to probe into the nature of suggested similarities and differences based on observed teacher behaviors. Questionnaires were evaluated and analyzed with descriptive statistics and mean responses for specific questions are reported.

In summary, Winter quarter research consisted of eight classroom observations of the same teachers teaching the same lessons observed in the Autumn quarter, one teacher interview, three short student questionnaires, and one comprehensive student questionnaire.

The entire research project was comprised of a total of eighty classroom observations involving five teachers, twenty teacher interviews, five written teacher questionnaires, and approximately 1000 completed student questionnaires.
3.7 Qualitative Data Analysis Procedures

The purpose of the qualitative data analysis was to be able to describe:

- The patterns and behaviors that indicated washback on the part of the teacher,
- The reasons that teachers gave for adopting instructional behaviors that may have been related to the test.
- The teacher washback behaviors that may have affected students’ behaviors.

For example, differences in student behaviors (noted in observations) included students frequently engaging in activities that demanded language skills such as reading, writing, speaking, and listening in Spanish, frequent use of Spanish as the primary language of communication, and decreased anxiety and frustration with regard to class activities that demanded language skills.

Qualitative data from teacher interviews, teacher questionnaires, and classroom observations was coded, evaluated, and analyzed for four main purposes:

- To identify changes in teacher behaviors observed by the researcher as a result of increased knowledge of the test.
- To identify changes in teacher behaviors self reported by teachers as a result of increased knowledge of the test.
- To identify teacher beliefs that appeared to influence teacher’s behaviors.
- To identify patterns in the behaviors of the teachers and students observed by the researcher throughout the quarter as a result of their increased knowledge of the test.

The purpose of the quantitative data analysis resulting in descriptive statistics was to further lend support to the descriptions of the patterns and behaviors that suggested
washback on the part of the teacher, and those behaviors that suggested washback on the part of the student. This data also provided further insight as to the reasons behind the teachers’ beliefs and instructional behaviors.

The quantitative data from the teacher and student questionnaires was organized and analyzed using Microsoft Excel®, and then enhanced using the statistical program Prism 4®, version for Macintosh. The reason for using Prism® is that it is intuitive, dedicated Apple® software, and produces concise and straightforward results, with customizable graphs and charts to elaborate results for the intended audience.

For the teacher questionnaire, several analyses were conducted: descriptive statistics for each question, a teacher mean for each question, a teacher mean for the questionnaire, and an overall mean for the questionnaire. The overall means were most effective at indicating the questions that received the highest responses from the teachers. The questions with the highest responses, and those related to test issues, revealed information that helped to describe some of the behaviors that were seen in the classroom observations. These results are reported in this study.

The self reported student study habits from the questionnaires that followed each test were analyzed in Microsoft Excel®. The study times for each test from each quarter were tallied. The totals were then compared to determine whether or not students’ study habits appeared to be linked to the differences in teacher behaviors that occurred as a result of the increased knowledge of the testing program as observed by the researcher.

The student responses on the Spanish 104 final exam questionnaire were organized and analyzed with Microsoft Excel®. Results of the students’ responses to questions focused specifically on tests were compared for individual teachers between
quarters to determine whether differences were suggestive of the observable differences in classroom behaviors of teachers as recorded by the researcher.

The primary purpose of this study was to focus on teacher data gathered via observations, interviews, and a questionnaire. This data provided information about teacher washback behaviors as they occurred in the low stakes testing environment of an intermediate Spanish foreign language classroom. Self reported data from all student questionnaires was secondary to this study. The student information was gathered to investigate possible trends that require further study, and to show the need for future studies that focus on student achievement as a potential means to measure and further describe washback in the intermediate foreign language classroom.

3.8 Data Reporting

The data gathered and analyzed for this study that pertains directly to the teachers in the study is reported in a “multiple- or collective case study” format (Fraenkel and Wallen, 2003, p. 439). This type of qualitative research is further described by Stake.

With even less intrinsic interest in one particular case, a researcher may jointly study a number of cases in order to investigate a phenomenon, population, or general condition. …It is instrumental study extended to several cases. Individual cases in the collection may or may not be known in advance to manifest some common characteristic. They may be similar or dissimilar, redundancy and variety each important. They are chosen because it is believed that understanding them will lead to better understanding, perhaps better theorizing, about still a larger collection of cases. (2000, p. 437)

By reporting the data in this manner, the reader is able to follow the study participants through the course of the research, linking each of the data gathering methods in progression (observations, interviews, and questionnaires) to the specific teacher. The result allowed the researcher to make comparisons between cases (the individual
teachers), even in instances where case comparisons appeared to result in contradictions (Stake, 2000, p. 444).
CHAPTER 4

RESULTS

4.1 Introduction

Due to the quantity of data gathered for this study, this section is organized around the different participants in the study and the kinds of data gathered. Each teacher is presented in a case study format and the Teacher Data that includes: Observations, Interviews, and Teacher Questionnaires. The Student Data includes: Student Questionnaires from Chapter Tests and the Final Exam. The Student Data as it pertains to each particular teacher is presented separately. Whenever feasible, data is organized into charts or tables, or when pertinent to the project, but too lengthy for inclusion as a chart or table, the data is presented in an Appendix. This format is to present the information in a coherent and logical manner, illustrating the described behaviors.

4.2 Review of Testing Program in Spanish 104

As described in Chapter 3, the tests that constituted the major focus of this study are considered low stakes tests, meaning the consequences of the tests are not used to make life changing decisions. Regardless of the stakes associated with these tests, they are still important to students’ success in Spanish 104 as they contribute to 55% of the final average (25% chapter tests, and 30% final exam). The other consideration
mentioned in Chapter 3 is a grade of 64 or lower in the Department of Spanish and Portuguese is considered failing.

Instructors teaching Spanish 104 for the first time in the Autumn quarter had all previously taught courses in the elementary level program. To reiterate the protocol for testing practices at this level, the point distribution for exams from Spanish 103 and Spanish 104 is included in Table 4.1. It should be noted that all exams in the elementary level (Spanish 101-103) follow a similar protocol as that listed for Spanish 103.

<table>
<thead>
<tr>
<th>Spanish 103</th>
<th>Points</th>
<th>Spanish 104</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comprehension</td>
<td>20</td>
<td>Listening comprehension</td>
<td>17</td>
</tr>
<tr>
<td>Dictation</td>
<td>10</td>
<td>Reading comprehension</td>
<td>18</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>20</td>
<td>Textbook reading questions</td>
<td>33</td>
</tr>
<tr>
<td>Structure and grammar</td>
<td>35</td>
<td>Language and grammar</td>
<td>32</td>
</tr>
<tr>
<td>Structure embedded in writing</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Culture</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL POINTS</strong></td>
<td><strong>100</strong></td>
<td><strong>TOTAL POINTS</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 4.1 Description of Spanish 103 and Spanish 104 test and point distribution.

The information in the table shows that for both Spanish 103 and Spanish 104 the students must complete a listening and reading comprehension section, which is valued about the same for each level (~ 40%). After that is where the tests begin show noticeable differences. Approximately 47% of the Spanish 103 test is focused specifically on grammar. In contrast, approximately 32% of the Spanish 104 test is focused specifically on grammar, and 68% is focused directly on skills (listening, reading, writing (about literary selections).
To further review the information presented in Chapter 3, the listening and reading comprehension sections of the Spanish 103 and Spanish 104 tests require the students to listen to a new auditory selection and answer questions in English based on what is heard and understood. They also read a new passage and answer comprehension questions in English. In Spanish 104, students must answer questions in Spanish about the required textbook readings from the chapters discussed in class. This section proves to be especially demanding for students, as they must answer the questions from recall without access to the stories. Typical questions written entirely in Spanish include tasks like explain the story, explain literary aspects of the story such as irony, describe elements of surprise, summarize the story, and/or describe the characters and their role in the story.

Although the tests from Spanish 103 and Spanish 104 do share similarities, the data gathered indicates that teachers and students view them as considerably different. For example, students are not accustomed to the quantity of target language (Spanish) writing required. The test also takes longer for the students to complete because of the increased writing. Additionally, the increased focused on skills presents new challenges to students as they learn how to best prepare for these tests.

For teachers, the increased focus on class readings presents instructional challenges. Teachers teaching Spanish 104 for the first time have not had to focus instruction as intently on reading and writing, and many teachers are learning how to do this best. Also, Test 1 remains largely unknown to them because teachers have never administered or corrected a Spanish 104 test before. Until teachers become more familiar with the testing program and its expectations, much of what teachers do to prepare
students for the tests (instructional behaviors and class activities) is based on personal opinion and/or past experience. Once teachers are familiar with the tests, although their instructional decisions are still personal, they reflect the demands of the test in a more obvious way.

Another aspect of the testing program that must be considered is the access that students may have to previous exams. Once all class members have taken a test, the tests are passed back to the students to keep. In theory this would seem to present a scenario whereby it is easy to obtain copies of exams to study. However, as a participant observer in 80 class sessions, and having had contact with more than 200 students, the researcher did not hear students discussing how easy they thought a test had been, or how they knew exactly what to expect on the test. Nor did students boast about high test scores. Rather, the researcher overhead, or was told directly by students and teachers (either in conversation or on questionnaires), how different the tests were from other tests the students had taken, how long the tests were, how low grades were, etc. It should not be assumed that all teachers and students shared these views, and certainly there were grades that were high (between 90 and 100) on tests, although these grades did not represent the majority of any one class. Therefore, it would appear to be inaccurate to assume that because a test accessibility issue exists, that everyone takes advantage of it.

4.3 Participant Profiles

The teachers who participated in this study were all teaching in the Department of Spanish and Portuguese at The Ohio State University. The following profiles provide background information about the teachers. The information is helpful to know in order to better understand how each participant manifested teacher behaviors and washback
behaviors in her Spanish language classroom. The information also provides insight into other factors with regard to teaching and testing that each contributed to the study. Pseudonyms have been utilized to protect the identity of the teachers.

Anne had been a teacher of elementary level courses in the Department of Spanish and Portuguese for the past two years. She began teaching in the department as a Lecturer immediately upon graduating from Ohio State with a Bachelor’s degree in Portuguese. After teaching for one year, she entered the Master’s Program at Ohio State in the school of Education: Teaching and Learning, Foreign and Second Language Education. While pursuing her Master’s degree, she continued to teach elementary level courses as a Graduate Teaching Assistant. At the start of the study, she had just finished her Master’s degree and began teaching as a Lecturer at the elementary and intermediate level.

Beth had been a teacher in the Department of Spanish and Portuguese for the past two years. She was a full time student pursuing her Master’s degree in Spanish Literature. While pursuing her Master’s degree she taught as a Graduate Teaching Assistant in the elementary level program. Upon conclusion of the study, she returned to teaching at the elementary level. She graduated with her Master’s degree Spring quarter, and planned to pursue the possibility of continuing to teach.

Cathy had been a teacher in the Department of Spanish and Portuguese for the past three years. She was a full time student pursuing her Doctorate in Spanish Literature. Prior to coming to Ohio State, she worked as a teacher at both the college and secondary school levels for a total of nine years teaching experience. Cathy was very forthcoming about the influence that the educational researcher Madeline Hunter has had
on her teaching; she made numerous references to her during interviews and even in class sessions. While pursuing her Doctorate, she taught in the department as a Graduate Teaching Assistant teaching elementary courses. During her participation in the study, she taught in the intermediate level program, and continued to teach at this level after the study was completed.

Diane had been a teacher in the Department of Spanish and Portuguese for the past two years. She worked a Lecturer teaching two to three courses per quarter at the elementary level. She had a Master’s Degree in Education, from the School of Teaching and Learning, Foreign and Second Language Education at Ohio State. Her primary foreign language is French. Her prior teaching experience consisted of a total of fourteen years of teaching at both the college and secondary levels. While participating in this study, she taught at the elementary and intermediate levels (Spanish 103 and Spanish 104). At the conclusion of the study, she was no longer teaching in the department.

Elena had been a teacher in the Department of Spanish and Portuguese for the past two years. She was a full time student pursuing her Master’s degree in Spanish Literature. Prior to coming to Ohio State, she worked as a teacher at the secondary school level outside of the United States for a total of twelve years teaching experience. While pursuing her Master’s degree, she taught in the department as a Graduate Teaching Assistant teaching elementary courses. While participating in the study, she taught Spanish 104, and continued to teach at this level after the study was completed. During the Spring quarter, she attained her Master’s degree and was subsequently admitted to the Doctoral program in Spanish Literature. During the summer quarter, she was assigned a 400 level course.
4.4 Overview of Teacher Data Collected

4.4.1 Classroom Observations: Autumn Quarter 2004, and Winter Quarter 2005

To review, a total of 80 classroom observations were conducted during the course of the research project: 40 in the Autumn quarter (8 for each of the 5 participants) and 40 in the Winter quarter (8 for each of the 5 participants). The classes observed were either directly related to the upcoming test (most often with the lesson focusing on one of the readings from the text De paseo that would be included on the test), or not directly related to the exam, usually focusing on either a cultural aspect, vocabulary, or speaking activities. To determine which classes to observe, the researcher consulted the calendar for Spanish 104. As earlier described, the Spanish 104 Supervisor sets the calendar prior to the commencement of the quarter, and it details each class session with the corresponding daily homework for the entire quarter. Although instructors do not have to follow the calendar scrupulously, it does provide a useful and logical guide for progression through the material that is included in each chapter, and the material that will be presented to the students on the exam. However, it was documented that although the calendar was consulted in order to observe lessons specific to the tests, there were instructors who did not follow the daily calendar, and therefore class observations did not necessarily reflect what the calendar listed, nor did these observations necessarily reflect aspects that would be included on the test, especially during the Autumn quarter of the research.

During classroom observations the researcher recorded everything that happened during each classroom session, including the time of all activities and instruction throughout the entire class session. Class observations were recorded using a model
adapted from the Saphier model for teaching, where the recording sheet was divided into two halves: one side for teacher talk, and one side for student talk. This allowed the researcher to record everything that occurred and was said during class by everyone in the room; and to see the distribution of student and teacher roles while accurately recording the time. Any other pertinent information to the observation such as information that was presented on the chalkboard or the overhead projector, whether it was homework, class work, or class activities, was also recorded. Additionally, if the instructor distributed an activity sheet, or a worksheet and she had sufficient copies, she would give the researcher a copy, eliminating the need to record the contents on the observation sheet. By recording the time of each and every activity in the class session, it was possible to calculate the division of instructional time for the eight sessions for each individual teacher. The class observation sheet is in Appendix A.

Descriptive details such as the number of students present in the room, and the number of males and females, appeared to have no effect on the teacher behavior for class lessons; therefore, this information was not included in any analysis of the classroom observations. The language proficiency level of the teachers was another factor that did not appear to influence the teacher behaviors in this study. Even though the language proficiency levels in both Spanish and English were markedly different among the participants, all of the teachers in this study were able to conduct their classes in the target language. The language ability did not influence whether or not teachers chose to incorporate certain concepts or topics.

To analyze the classroom observations, each observation was organized into the percentage of instructional time spent on a specific activity. A tally was completed for
each instructor for each observation for each quarter. The complete tallies for Autumn and Winter quarters are included in Appendix D due to the large quantity of the data. The result of these tallies show that clear teacher behaviors reflecting washback emerged. These behaviors arose from the teacher’s knowledge of the test, her understanding of the students’ needs with regard to the test, and her understanding of the nature of the intermediate language course. Other behavior patterns emerged that reflected developing teacher behaviors as a result of increased experience with the Spanish 104 course, and increased teaching experience. These washback and teacher behaviors included: teacher behaviors that focused on grammar, washback behaviors that focused on textbook based reading and writing activities, teacher behaviors that focused on past teaching and washback behaviors that focused on past testing experience, and teacher and washback behaviors that focused on general skill building in reading and writing.

During the Winter 2005 quarter, observations included the same five participants and the same class sessions according to the course calendar. It should be noted that one instructor taught her class during the evening, so instead of spending 48 minutes a day for 5 days, she spent 2 hours twice week. Therefore, some of the observations lessons from this instructor are longer than 48 minutes, in order to provide continuity to the observation.

As mentioned, not all teachers followed the calendar closely, and therefore some of the observation lessons contained different content than the first observation of the same lesson for the Autumn. In other words, although the calendar indicated the same particular lesson for the day of the observation, it was not uncommon during the Autumn quarter to find a completely different lesson being presented. However, during the
Winter quarter, all teachers seemed to follow the calendar more closely, and the lessons observed more closely reflected the content of the course calendar. Teachers still made adaptations allowing for personal adjustments to accommodate instructor generated projects and activities, but the adherence to the calendar revealed noticeable differences. Washback was more evident through the shift toward using more test-like activities in every classroom. The degree to which test-like activities encompassed class instruction varied according to instructor; however, there were clear and observable differences between the instructional behaviors exhibited by the participants in the Winter quarter as compared to those behaviors exhibited by the participants during the Autumn quarter. Many of the same types of behaviors were evident, which is not surprising given that the curriculum was the same. However, the time afforded to these behaviors, and the instructional decisions based on these behaviors, were much more test centered for the majority of the participants, thus indicating varying levels of washback. Also, according to observations and interviews, teachers felt more comfortable teaching the course the second time during the Winter quarter. Not only were they more familiar with the testing program and the demands that it would place on both instructor and students, but they were also more familiar with the course curriculum, and with the types of questions and concerns that students would present.

4.4.2 Teacher Interviews

During the Autumn quarter of the research, three interviews were conducted with each teacher. These interviews took place after Test 1, after Test 2, and either before or after the Spanish 104 Final Exam. During the Winter quarter, one interview was conducted with each teacher. This interview took place either before or just after the
Spanish 104 Final Exam. The purpose of all of the interviews was to gain further insight into the teacher behaviors observed in the classroom, and into the teacher’s perceptions of how the tests influenced her teaching. The second quarter interview was also to gain further information about self-reported changes in teaching. Each interview was tape recorded and then transcribed. The transcriptions were then analyzed to identify patterns in teacher behaviors with regard to class instruction, to mark contradictions between what had been observed in the classroom and what the teacher indicated during the interview, and to fully describe the self reported changes observed and reported by the teachers.

4.4.3 Instructor Questionnaire and Responses

During the first quarter, each teacher was given a questionnaire to complete after the administration of Test 2 and prior to the administration of Test 3. The purpose of the questionnaire was to determine the instructors’ perceptions of the demands placed on her as a result of teaching Spanish 104, and teachers’ attitudes toward, and perceptions of, language learning. After the observations and interviews it was clear that teachers had specific ideas about how to focus their instruction. These ideas came from the tests and course demands, and also from the teacher’s beliefs about language teaching and learning. In order to gain more information regarding these beliefs that affected instruction, teachers were asked to respond to the following series of questions:

1. What kind of extra work or pressure, if any, do you think the Spanish 104 tests have put on you in your teaching?

2. What are the most significant changes you have made in your teaching in the context of the different exam format for the 104 tests?
3. What do you find the most difficult or challenging aspects of teaching Spanish?

4. What are the learning strategies you would recommend to your students in the context of the transition between the elementary language program and the intermediate language program?

5. What types of activities do you believe are most effective in language learning?

6. What do you believe to be the major aims for studying and learning Spanish at the Ohio State University?

7. In what ways would you like to motivate your students to learn Spanish?

8. The factors that most influence your teaching are:

   [ ] Professional training
   [ ] Academic seminars or workshops
   [ ] Teaching experience and beliefs
   [ ] Teaching syllabus
   [ ] Past experience as a language learner
   [ ] The need to obtain satisfaction in teaching
   [ ] Textbook
   [ ] Exams
   [ ] Learners’ expectations
   [ ] Peers’ expectations
   [ ] Supervisors’ expectations
   [ ] Social expectations
For questions one through seven, teachers were asked to rate specific prompts as potential responses. They were also given space to add their own comments whenever appropriate. The prompts were chosen to reflect the possible perceptions and feelings of teachers teaching a new course for the first time as determined by the researcher, the Spanish 104 Supervisor, and the Spanish Language Program Director. For example, in question three, teachers were asked to comment on the aspects of teaching Spanish 104 that they found the most difficult by rating the following prompts: students’ current level of Spanish, class size, inadequate teaching resources, learning environment, limited teaching and learning aids and facilities, overly demanding workload, inadequate time for students to practice Spanish outside of the classroom. Teachers rated their answers on a Likert type scale arranged as follows: 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree. Teachers responded according to their beliefs and perceptions after having taught the course for approximately eight weeks. Because this questionnaire was distributed to the teachers close to the end of the Autumn quarter, the researcher did not have the opportunity to observe many of the self reported changes until the Winter quarter. A copy of the complete questionnaire is in Appendix B.

The questionnaires were analyzed using Microsoft Excel ® to determine the mean response for each teacher, the mean response for each question from all teachers, and the total mean response for the questionnaire. Individual teacher means were then graphed using Prism®. The results were analyzed in the following manner: the individual teacher mean for the entire questionnaire was reported to show the teacher’s overall response agreement with the items on the questionnaire. Analysis then explored the questions that
received higher than the teacher’s mean response, and those questions that focused specifically on tests.

4.5 Organization

This section presents the data gathered specific to each teacher. For each individual participant in the research, the Teacher Data (classroom observations, teacher interviews and the teacher questionnaire) has been organized into a section devoted entirely to that specific teacher. Section 4.6 presents the data gathered from Anne, Section 4.7 Beth, Section 4.8 Cathy, Section 4.9 Diane, and Section 4.10 Elena. Upon conclusion of these sections, summaries of the data components are provided. These include summaries of the Observations and the common themes, summaries of the Interviews and common themes, and a summary of the Teacher Questionnaire. The Student Data section follows the summaries. Results from the Student Questionnaires are in the Student Data section.

4.6 Anne

4.6.1 Observations

Autumn quarter observations from Anne’s classes were characterized by teaching behaviors that relied heavily on past teaching practices. Winter quarter teaching behaviors changed as a result of increased familiarity with all aspects of Spanish 104 course, and increased confidence. Consequently, Anne demonstrated more washback behaviors that directly reflected the Spanish 104 tests.

During the Autumn quarter, Anne exhibited behaviors that focused extensively on explicit grammar instruction with the students, and she frequently used activities that were recycled from teaching an elementary level course. Her behaviors were indicative
of those arising from past teaching and testing experiences. It was common to find her using grammar worksheets from elementary level courses that she had taught, and spending significant time during class sessions focusing on explicit grammar instruction. Sometimes these activities were explicit grammar drilling exercises, and other times they were activities like songs. She believed that the rules governing the uses of the preterite and imperfect tenses, and the uses of the subjunctive and indicative tenses, were of extreme importance to students. These tenses are presented and tested extensively at the Spanish 103 level. Although the tests in Spanish 104 do not focus as extensively on grammar, Anne seemed to still be teaching as if preparing students for the Spanish 103 testing program. In an Autumn interview, Anne commented on her focus on the grammar tenses:

Today we did a review of preterite and imperfect because that’s the grammar section for this chapter. Um, but like I gave them a mnemonic SAFE WATERS. SAFE has to do with when you would use the preterite tense, and WATERS is when you would use the imperfect. … I had a story that we read through and um, for anything that they thought the verb in capitals should be in imperfect they did this (starts clapping legs). And anything for preterite they clapped once.

To enhance her direct grammar instruction, Anne often used comprehensive review sheets or special activities not contextualized to Spanish 104. These review sheets typically included some type of acronym or mnemonic in order to help students remember the information presented. For example, for the common uses of the subjunctive tense, a popular mnemonic is WEIRDO, with each letter representing a reason to use the subjunctive tense in Spanish. The “trick” to help the students remember if they need to use the tense in a particular situation or not is as follows: W-wants/wishes, E-emotion, I-impersonal expressions, R-recommendations, D-doubts.
The review sheet also contained a list of the common verbs in each category with practice activities for each category. The worksheet was used with students in Anne’s class, and in other classes as well. Anne was happy to share her worksheet with other teachers.

Activities also focused on form and accuracy, the most common way to test those structures in Spanish 103. One special activity focused on grammar was used to help students with the formation rules of the imperfect subjunctive. Anne taught the students a chant, “kill-Ron and cheer”, which they shouted enthusiastically while she referred to them as “cheerleaders”. The purpose of the chant is to teach students how to form the tense properly. The “ron” part of the chant is the ending from the third person preterite form of the verb (ellos, ellas, Uds.-aron/-ieron) that must be dropped from a verb form. The “cheer” is represented by a form of “-ra” which is the appropriate ending to add for formation of this tense (-ra, -ras, -ra, -ramos, -aron) based on the subject. The uses of past activities, specific grammar materials, and test like activities from Spanish 103 indicate that Anne’s behaviors in the Autumn quarter were behaviors reflective of her past teaching experience, and some washback from the tests in Spanish 103.

During the Winter quarter, the teaching behaviors that focused on grammar became a less dominant focus of the classroom instruction. That is not to say that Anne discarded grammar instruction; however, it was less common to see her spend an entire class session focused on de-contextualized grammar. Anne used the textbook grammar activities more often, and/or she built upon those activities to generate her own activities. Overall, during the Winter quarter there was much less teacher behavior reflective of past teaching experience, and less demonstration of washback behaviors from Spanish 103 testing practices. Although there was still direct grammar instruction, for each entire
class session that was focused on grammar during Winter quarter, Anne consequently devoted more than half of remaining class sessions predominantly, or entirely to reading activities and/or explicit current test activities.

When it came to teaching reading and writing during the Autumn quarter, Anne spent little time on the reading and writing activities that focused on the textbook readings. She mostly assigned the readings and the *Preguntas de Orientación* but spent little class time discussing them. There was also limited use of activities that expanded upon the assignments and not much focus on instruction to develop students’ reading and writing skills.

Increased writing and story analysis are direct Spanish 104 test activities. During the Winter quarter, students in Anne’s class often engaged in reading activities that included pre-, during-, and post-reading activities, and literary analysis of the readings from the chapter. It was more common for Anne to ask her students to write summaries of the stories they had read, and to break students into groups and have each group be responsible for a summary and explanation of their part of the story. The students also practiced writing, and talked about how to write well in Spanish. Students shared their writing with the rest of the class either on the chalkboard or on an overhead transparency, and with the teacher.

Discussion and interpretation of the textbook readings became an important part of the behaviors that focused on reading. Anne was effective at being able to incorporate her own questions along with the *Preguntas de Orientación* in order to further probe the students’ understanding, and their ability to communicate this understanding. She was
also able to relate the pre-reading activities to the students’ context, and therefore able to successfully scaffold the readings for the students.

A common test activity was to ask students to write a summary of one of the stories they read; however, this was not a part of every test. During the Winter quarter, students in Anne’s class frequently engaged in activities in and out of class (homework) that encouraged the development of their reading and writing skills in general. This became a common activity that Anne frequently used so that students were no longer overwhelmed when asked to write 10 to 12, or even 15 sentences about a story in class, or for homework.

To promote reading skills, Anne incorporated increased time for teaching reading strategies to students. She focused on the importance of reading the title, and the information that the title alone could provide for the reader prior to initiating reading an unknown selection. She also stressed the importance of the Pequeño Diccionario in the textbook, a small glossary provided prior to the reading that defined Spanish words in Spanish, and contained simple illustrations. This glossary provided the meaning of the word within the context of the textbook reading. By focusing on these strategies, Anne’s students were more apt to refer to these resources prior to reading even on their own. Because Anne stressed the importance of titles and their meanings, the vocabulary, and the story organization, her students were more apt to use these tools when reading. This in turn helped to prepare her students for success on the listening and reading comprehension sections of the test. During Winter quarter Anne was also observed following the calendar more closely and focusing more completely on the textbook
reading and accompanying activities throughout the entire class sessions as specified on the calendar.

Consequently, in the Winter quarter skill development became a focus in her classes. The teacher and the students began to view reading as an integral part of the intermediate classroom, and each began to realize that grammar was no longer the primary focus of the curriculum at this level. Students were more engaged in conversations about the textbook readings both with their teacher and with the other students in the class. They appeared to ask fewer clarification questions about the stories, and about subsequent activities that focused on post-reading activities. For example, in Chapter 3 *Presagios*, an authentic piece of literature, was a required reading. This is a particularly challenging text for the students, combining reality with dream sequences. In Anne’s class during the Winter quarter, students were observed working in groups outlining the story and organizing it into the dreams of the narrator and his actual reality. In their groups, they discussed the story and their questions. Anne led a discussion reviewing the entire sequence, and followed with a description of the literary terms used in the short selection. The groups then read the story together and acted it out. These behaviors exhibited by Anne indicated washback as they engaged students in activities that helped them to prepare directly for all sections of the test that demanded skills.

One example that demonstrated the changes in her behaviors is the lesson when Anne conducted a *Jeopardy®* review prior to an exam. The Jeopardy® game focused entirely on the stories that would be on the test the following day. There was no mention of grammar, nor were there any grammar categories in the game. The students in her class wrote the questions about the stories used in the review. They had handed in the
questions the day before to Anne, she made necessary corrections, and then developed the game. At the end of the review class, she ended by saying to the class “now you know what you need to study [for the test].” The importance of reading and writing both as a significant portion of the test, and a crucial skill in the language class was something that Anne understood. She realized that skills constituted a major portion of the testing demands, and she made sure that she incorporated activities to develop these skills into her teaching repertoire.

The washback behaviors in the Autumn that seemed to focus on Anne’s past test experience (grammatical form and accuracy), had shifted to present test experience during the Winter quarter. Her washback behaviors in Winter quarter included using direct test activities such as discussion of and activities for new readings, listening comprehension activities, and video viewing activities that focused on listening comprehension. These behaviors indicating Spanish 104 test washback demonstrated that there was a greater differentiation for instruction between what Anne previously believed students “should” and “need” to know based on her prior teaching and knowledge of Spanish 103 testing, and what she acknowledged that her students needed to know for success on the Spanish 104 tests. Even though Anne did not minimize the importance she gave to grammatical accuracy, she was able to see the importance of the reading and writing activities within the context of the 104 course.

4.6.2 Interviews

Anne’s comments during the interviews mostly reflected what was observed in her classes. Discrepancies between her perceptions, the researcher’s observations, and the student’s perceptions were more apparent during the Autumn quarter. The researcher
noted that particularly during the first two interviews, Anne seemed concerned with answering in a way that she thought was appropriate or expected. For example, at times it appeared as though she was trying to impress the researcher with her answers, but instead appeared unsure of herself because the question would be unanswered, or the answer would be unrelated to the question. Some answers also reflected her tendency to try and be creative in her classes, which was confirmed in observations. Although Anne had many creative ideas, it took ten weeks for her to realize that the amount of creativity did not necessarily help to prepare students for the demands of the course.

Although the Autumn quarter observations revealed a fairly intense focus on grammar in Anne’s classes, her comments during the interviews indicated that she did not perceive her classes to have this focus. She said:

….it wasn’t just a matter of like, ok, I have to write this down correctly and say it correctly, but like, actually engaging in the message beyond the grammar. Because my understanding of this course, is that it’s not about the grammar, it’s about the skills; the reading, writing, listening, speaking skills, and, so, um, we’ve hit a lot…ah my approach, um, it’s been communicative, but a lot of it’s been strategic.

Although Anne perceived that her classes focused a lot on the readings, she acknowledged a shift to more grammar activities after the first two tests:

After the first exam, I realized that not doing much in the way of explicit grammar review was detrimental. They apparently weren’t doing it on their own…I mean granted it was only like 35% of the points, but that’s the difference between passing and failing. So um, I have been spending more time on the grammar, giving them handouts that like basically summarize it in different forms. …in my opinion, we do a lot of activities in class, we do a lot of talking about the readings, we’ve done a significant amount, at least you know, the last 7 weeks, we did a lot more grammar…and examples, and putting it on the board.

Following the second part of this comment, Anne actually turned off the tape recorder during our interview to talk with me “off the record”. She was very concerned about how
some students had responded to the question asking whether or not they felt that the teacher prepared them adequately for the exams. Although her perception was that she did a good job, 33% of her students responded with a “Not significantly (2)”, or a “Not at all (1)” in terms of clearly communicating expectations for the test. This supported evidence gathered in observations that during the Autumn quarter Anne struggled to find the balance between her emerging teaching repertoire that was influenced by her experience in Spanish 103, and meeting the expectations and demands of Spanish 104 and the assessments.

At the end of the Autumn quarter, it was obvious that Anne’s first time teaching Spanish 104 had left some lasting impressions on her teaching plan. These impressions served to influence her instructional behaviors for the subsequent quarter. Anne commented about getting ready to teach for the second time saying:

I know that because I understand the text and the workbook better, I will explain things up front. Like Day 1, I’m only going to spend like maybe 15 minutes on the ‘go and get to know each other thing’, and I’m going…to explain…[the layout of the workbook] and talking about just like the structure of the chapters…

Her Winter quarter classes included more emphasis on reading and writing, specifically with activities that reflected direct Spanish 104 test activities. She mentioned:

For every reading they had a follow up assignment. Either they had to come up with their ten original questions, which I did sometimes last quarter, um, or they had to do a ten line double spaced summary…I think I pushed them a lot further than what any exam would ask of them in terms of how deeply we analyzed, and how detailed [we were]… I think sometimes when they get to the exams they feel that it’s easier than what I’ve made them do in class.

When she was questioned during the Winter quarter about differences she observed in her teaching, she made specific reference to the shift from grammar that dominated much of the instruction during the Autumn quarter.
… I think last quarter I focused more on grammar, and I think that I touched on grammar, and like I definitely made them work it and stuff, but I think I gave a lot more. I gave usually about two days, or a day and a half to each of the readings. Um, and so that was probably the biggest difference because up until last quarter, I was a grammar teacher – that’s basically what 101-103 is- and so I have to switch gears. I was reading with them for the first time like the night before or two nights before them, so I wasn’t necessarily prepared with the kind of questions or problems they were going to have with it, and um, I guess I didn’t understand how much I needed to help scaffold the reading. … Since it was something that I had never really worked with, and didn’t have much experience with, I didn’t really know how to do as well.

When asked how she helped to prepare her students for the tests during the Winter quarter in comparison to the Autumn quarter, Anne revealed her use of direct test-like activities:

… I guess I knew about the tests so I started going on the stories more. What kind of topics would be asked of them, and the production of, and the writing for those stories.
… I was very explicit. I’m like you need to know… the readings. If you know nothing else, you have to know the answers to the Preguntas de Orientación because I was like you know, that is a given, that’s basically an outline, a lot of times pinpointing some of the very important stuff that they want you to get out of it. Um, I never explicitly said that last quarter, I guess I always just assumed that it was self-evident, but a lot of thing that we assume… anyways, so I explicitly said that.

Anne reported that she felt more comfortable teaching the class for the second time during the Winter quarter, as she said:

… well last quarter, the whole thing I was learning how things were working, what they were looking for and stuff like that, and so last quarter prepared me for this quarter. And um, since I had prepared a lot of activities, I could expound and reinvent other different things…

Anne further added, “… I think I’m a much better instructor this quarter around, but I think it’s just being familiar with it, usually the second time goes a lot better.” The familiarity with the course content and expectations affected Anne’s confidence and her choice of instructional activities as is expected. But it was also apparent that
increased knowledge of the tests in Spanish 104 influenced her teaching as reflected in her choice of more direct test-like activities during class time, and a greater focus on test preparation during the second teaching. Anne exhibited washback behaviors from the Spanish 104 during the Winter quarter more than during the Autumn quarter.

4.6.3 Anne’s Questionnaire Results

For Anne’s questionnaire, her average mean response was 3.86, a response rate that is above neutral and slightly below agree for most questions. Some interesting details were revealed in her responses that supported what was observed in her classes. For example, in response to question one about extra work and pressure placed on her from the tests in Spanish 104, the prompts she ranked in the 4 and 5 range of the scale included: following a new syllabus, revising the existing materials that she had, meeting new challenges in teaching, and organizing more focused activities that reflect exam activities. In question two she responded that one of the most significant changes that she had made to her teaching was emphasizing the integration of skills. Although the questionnaire was distributed toward the end of the Autumn quarter, this change was readily visible in her classroom during the Winter quarter.
As Figure 4.1 indicates, the two questions that received the highest response (indicating the most agreement) from Anne were questions four and five: learning strategies that she would recommend to her students, and effective activities for language learning. For the learning strategies, Anne felt that students should emphasize those for speaking, listening, and reading; expose themselves to Spanish realia outside of the classroom; create opportunities to use Spanish in daily life and communicate more in Spanish. She underlined the prompt emphasizing the importance of the transition from passive learning to active learning. The focus on skills was becoming a cornerstone of her beliefs regarding language learning, and this was apparent in her teaching during the Winter quarter.

Interestingly, Anne also commented that she was neutral with regard to the role of authentic materials in language learning, and neutral with regard to the influence of exams on her teaching. She indicated that professional training, teaching experience and
beliefs, experience as a language learner, and learner’s expectations were much more important as influences to her teaching, than was teaching to the tests. Her responses on the questionnaire supported the teacher behaviors observed during the Autumn quarter; however, the responses seemed to contradict her behaviors observed during the Winter quarter, when she spent considerably more time on direct test-like activities, thereby demonstrating washback from Spanish 104 tests.

4.7 Beth

4.7.1 Observations

Beth’s classroom observations were characterized by frequent use of past teaching materials, heavy reliance on the textbook, and limited teaching and testing experience. Her class activities were often modeled on information from other textbooks, and were sometimes borrowed from other teachers. Beth appeared to be unsure how to get students to do what she wanted them to do during class sessions. During the Autumn quarter, Beth exhibited teacher behaviors that focused extensively on explicit grammar instruction with the students, and she frequently used activities that were recycled from her teaching at the elementary level.

In the Winter quarter, Beth did not seem to demonstrate a notable difference in her focus on grammar behaviors. She again employed many of the grammar activities from prior teaching that did not relate to the theme of the chapter or the readings in the chapter. For example, although she did use more grammar practice activities from the textbook, she also still employed pair work activities (A/B information gap activities), activities from previous courses that she had instructed, and overhead transparencies that were clearly contextualized within themes of a different text. One overhead activity that
Beth used was about a trip to Spain, and in another activity she asked students to identify their preferences with regard to school and activities. These themes were part of the elementary level curriculum that uses the textbook Amistades, and not included in De paseo chapters that are part of the Spanish 104 course curriculum.

There were also instances in Beth’s class where she used worksheets that were obviously from a different text altogether, or borrowed from someone else. Although during the Winter quarter Beth did exhibit teacher behaviors that used more grammar activities from the De paseo text than she did in the Autumn quarter, she continued to supplement book activities with worksheets or overheads that she had used in other classes. Beth’s teacher behaviors during the Winter quarter illustrated that grammar was still a predominant focus for instruction in her Spanish 104 class, as the percentage of instructional time devoted to these activities remained fairly constant.

Following the calendar more closely was a characteristic of the other four teachers in the study during the Winter quarter observations. However, Beth’s did not seem to follow the calendar any more closely than she had during the Autumn quarter. For example, during a class observation of a test impacting lesson in the Winter, Beth conducted an activity that was modeled after the textbook reading (an interview with Gloria Estefan) where students had to interview each other and then present their interview to the class. This was followed by a grammar activity that was related in vocabulary and theme to the reading selection Nuestra Gloria that Beth had completed in class the day before. Observations and interviews revealed that teachers were challenged to follow the calendar in terms of keeping up with the materials offered for each class session, and it was rare to find teachers ahead of the published calendar. A possible
reason for Beth being ahead of the calendar lesson is limited instructional time devoted to this reading and the subsequent activities, which in turn resulted in limited instructional time on test activities.

Characteristics of Beth’s limited teaching repertoire were more apparent during the Winter quarter, especially with regard to teaching reading and writing. Although the other teachers were not satisfied that students understood the story based solely upon the answers to the *Preguntas de Orientación*, Beth still used these questions as her primary teaching tool for reading. She found it particularly challenging to manipulate the activities from the text. Although she always used the *Preguntas de Orientación*, she rarely discussed these questions with the students. She would ask the students to compare their answers with a partner, or in groups, and then after about 5-10 minutes, she would ask if there were any concerns or problems (“¿problemas o preguntas?”) about the questions. If there were none, and students hardly asked, she would continue on with the lesson and her instruction. Sometimes Beth would move to a post-reading activity, although there had been no during-reading activity, or at other times she would move to a grammar activity. She assumed that because no questions were asked, there were no questions or comprehension problems, and she didn’t need to discuss the material further with the class. Beth specifically addressed this issue in interviews. Moreover, whenever students were assigned to read a story for homework, she characteristically began class the next day with a quiz about the reading. The quiz was short, typically 10 questions, and usually True or False. She would give the students approximately 5 minutes to complete the quiz. No discussion of the reading was ever held prior to, or after the quiz. Beth would move into the next activity, which was often the aforementioned *Preguntas*
de Orientación. Her washback behaviors reflecting test expectations that focused on using textbook reading and writing were limited. Beth did use other activities in the text, but she was not as adept as other teachers at manipulating the given activities to the specific context, or to her individual class.

During observations of Beth, it was not evident, nor did students exhibit behavior that suggested that she spent focused time on developing reading and writing skills. When asked to read or write spontaneously, the students in her class were clearly uncomfortable with activities that demanded the use of these skills. When asked to use their skills for an activity, students asked frequent questions of the instructor after an assignment or task had been introduced or explained, and/or they often asked the researcher questions such as “how would I say…?”, “are we supposed to do this?...”. Sometimes in groups, one student would take charge and complete the activity while the others express concern for their lack of skills in writing, or reading, or their low confidence in their comprehension level.

Past testing experience from Spanish 103 seemed to play a role in Beth’s instructional behaviors. Like Anne, Beth, also believed that the rules governing the use of the preterite and imperfect tenses, and the uses of the subjunctive and indicative tenses to be of extreme importance to students. Because she had had experience teaching and testing in Spanish 103, her activities in Spanish 104 suggested washback from the Spanish 103 tests. Beth borrowed review sheets, used the same mnemonics, and spent considerable instructional time focused on explicit grammar. Her class sessions during both the Autumn and Winter quarters often focused on direct grammar instruction and practice.
During the Winter quarter, in contrast to other teachers, Beth did not balance out her grammar instruction with reading and writing activities, nor did she demonstrate many major instructional changes from the Autumn quarter observations. She did not appear to demonstrate increased washback behaviors from the Spanish 104 tests. An example of this comes from one of the class sessions the day prior to an exam. Beth spent no time on discussion of the textbook readings included on the exam. Rather she focused entirely on the grammatical points included in the chapters. This was a sharp contrast to the Jeopardy® review (conducted by Anne) mentioned earlier that focused entirely on the textbook readings. Anne used the review time to focus on the readings acknowledging the emphasis given to the readings on the exams, whereas Beth focused on grammar. Beth still viewed grammar as the most important aspect to know for the test, as she believed that this was the part of the test that would be most problematic for students.

4.7.2 Interviews

During her interviews, Beth provided support for what the researcher had observed. As a relatively inexperienced teacher, she had a limited teaching repertoire, and she faced many of the challenges that confront new teachers. Her greatest were getting students to do what she wanted them to do, relying heavily on materials from past course and on the textbook, and gauging student comprehension during instruction.

Beth was excited as she began teaching Spanish 104. She commented:

…I feel like with these students in particular, it’s going to be a lot easier to sort of bypass the sort of plans I made in 102 and 103, where we’re not really gonna have to do the little info gaps and the little fill in the blank as much, so I feel like I don’t have to prepare like a whole presentation to present the verbs or anything.
…and as far as the reading, I told them we’re not going line by line at all, and to sort of ensure that, I’ve been giving little quizzes before we go over the reading activities, so that I know that they’ve read, and, it’s so far, it’s worked out well. It looks like they’re doing the _Preguntas de Orientación_ and….so my approach is more, let’s have fun with the stuff in class vs. actually going through the stuff, so it requires a lot of them outside of class.

Beth seemed to be under the impression that perhaps the class would demand less from her in some ways because it was a more advanced class, and the students were going to be focusing on different types of activities. However, interviews with Beth after the first weeks of class and Test 1, revealed a different perspective. When asked how she had changed her teaching since the first exam, she said:

I know that for the grammar, I included a lot more of the activities that I did in 103 and 102. As far as like, more grammar based activities, I am still doing the ‘entrevistas’ and stuff like that, but also doing more of the little info gap things and stuff…and a lot of times I’m making sure, before we even start doing activities, that they do remember the grammar, instead of just saying, “ok, let’s just go over this.

She further commented on the shift in instructional emphasis after Test 1, which in her classroom was a shift toward more grammar instruction and activities:

…I’ve just cut out a little more of the straight conversation stuff. They’re still included but I’m making more room for the info gap activities, and stuff like the fill in the blanks to just make sure that they have basic knowledge before we do start using things in conversation.

Beth also revealed after Test 1 and Test 2 that she believed the largest deficiencies in students’ skills in Spanish 104 were grammar skills:

…I thought that they were going to be able to do a lot more than they actually could, and they still had a lot more problems. They had a lot more problems, then I would have thought that they had….specifically the grammar, specifically the grammar. I mean after, I mean after I’ve even taught 103 and I know that they get so many tenses thrown at them in 103, and then in 104 we started with the present tense, they were confused, they, there were just like, wait a minute, they had to go back and think about it again.
In a comment made during the second Autumn interview, Beth revealed what was observed in her classes; her limited instructional time on test like activities, especially in the area of reading and writing, was affecting the students. She said:

I think they were kind of upset after the first test. Partly because I guess maybe they felt like it was overwhelming and I think they didn’t realize how important the stories were. They understood how important the grammar was, and I think for most of them it was a review that they weren’t too worried about. Even when I did tell them that the stories are gonna be on there, you should be able to name characters and talk about the stories, they still had no concept of how they were supposed to do that or whatever….

Beth reported that she was “trying to do more activities, with the reading that do make sure that they have a basic knowledge before we talk about the story, before we go a little farther”. Beth’s comments in the interview contradicted what was observed in class sessions. She spent limited time on the readings themselves, and on pre-, during-, and post-reading activities. She often considered students’ group comparison of the Preguntas de Orientación as the determining factor in whether or not students achieved reading comprehension. She rarely discussed the questions as a class before moving on to the next class activity. Asking “¿preguntas o problemas? [questions or problems?]” was sufficient before moving on.

At the end of the Autumn quarter, when Beth was asked what she believed the best way to help the students prepare for the test was, she responded:

Uh, I still don’t know. Because they ask me, and I say you know, you can think about the types of questions that have already been on your tests, but I don’t know what to tell you. And so, one of my students said, ‘so we should probably be prepared for the worst case scenario.’ She said, “what is the worst case scenario?” And I said, “well in my opinion it would probably be having to summarize the whole story. And she goes, “yeah, that’s the worst case scenario.” So that’s the only thing I feel like I can tell them.
Even though the first quarter of teaching this course had been completed, Beth still did not feel as though she was able to anticipate the demands of the tests, nor did she believe that she was capable of adequately preparing her students for these demands. She believed that much of successful test preparation was left to chance, and the amount of studying in which students themselves engaged.

During the Winter quarter interview Beth reported not making any significant changes that resulted in increased attention to reading and writing, and the classroom observations confirmed this. Her answer about changes that she had made to her teaching was vague. It also demonstrated that Beth had not been able to critically reflect on her teaching. She did not spend increased instructional time on reading and writing skills, and she continued to treat these skills in mostly the same manner that she treated them during the Autumn quarter. She said:

I think a lot of my changes are the same, I think a lot of what I, I’ve changed more since the first test of last quarter, that I have, I mean after that first test I changed a bunch of stuff after that. But since last quarter, I don’t think I really changed a lot.

When questioned as to the emphasis she gave to reviewing the textbook readings prior to an exam, Beth responded,

…for this last chapter, I don’t think I did…no we focused on grammar which a couple of students were disappointed on, but I think when we actually did those, the readings in class, I think that I tried to emphasize the same kind of things that I thought the test would, so it wasn’t exactly a review, but I just tried to be really thorough, knowing that we wouldn’t have time to review the stories, because, we would only have time for the last two days to talk about all the grammar.

Beth did not make the connection between a shift to reading and writing, and a link to developing skills for the students and preparing them for the demands of the testing program. Rather, she was focused on the importance of grammar for language learning
as an indicator of students’ progress. The fact that grammar was not the main focus of the present testing program, did not influence her instructional decisions. Beth continued to implement more and more grammatical activities and direct instruction throughout the Winter quarter in spite of the limited emphasis given to this area on the tests. She attempted to use the reading activities from the book, thus presumably acknowledging the focus that the test placed on reading; however, in spite of this, she appeared unable to manipulate the textbook activities sufficiently to direct and focus students’ attention to the skills of reading and writing. Her limited experience and teaching repertoire influenced her instructional decisions more than the tests.

4.7.3 Beth’s Questionnaire Results

For Beth’s questionnaire, the mean response was 3.58, which is slightly above neutral, and below agree for most questions. She indicated that more lesson preparation and organizing focused activities that reflected exam activities were the greatest demands placed on her in teaching Spanish 104. The most significant change that she reported to her teaching was encouraging more student participation in class; however, the researcher did not frequently observe this in her classes during the Winter quarter. Figure 4.2 illustrates Beth’s mean response to each question for the questionnaire.
The question that received the highest rate of agreement for Beth was number four. According to her, the learning strategies that she would most recommend to her students were the following: exposing themselves to Spanish realia outside of the classroom; expressing their opinion more frequently in class; emphasizing speaking; reading, and writing; learning to initiate questions; being more active in class participation; communicate more in Spanish; and transition from passive learning to active learning. Reviewing Beth’s responses, it appears that she made the connection between the importance of skill development and language learning, and that she perceived the importance of these skills on the tests. She even stated on her questionnaire that she would focus more on the areas of listening, reading, and writing, thus suggesting that she understood the need for students to be prepared in these areas for the testing program. However, Winter quarter observations revealed that Beth found it challenging to do this, as her instructional focus in these areas was limited.
Beth indicated that the important factors most influencing her teaching were teaching experience and beliefs, past experience as a language learner, the need to obtain satisfaction in teaching, the textbook and exams, and supervisors’ expectations. The fact that her experience was solely in the elementary program would support this as her teaching was observed to reflect elements of the elementary course and the elementary testing program, such as the increased focus of grammar. The importance that Beth attributed to her past experience as a language learner is interesting because she was never observed relating her own learning to the student’s learning in an attempt to facilitate language obstacles for the students. Because she is a Master’s student in Spanish Literature, she has real experience with learning to read and write in the target language. However, this was something that she did not overtly link to her own students’ learning as they attempted to grapple with these same skills, both in the class sessions, and in preparing for the tests. One would have expected to see Beth’s instructional behaviors reflect the factors that she indicated as most influential on her teaching, such as including more test like activities, and using her own language learning experiences to facilitate her students’ language learning and skill development. Instructional behaviors that focused on these aspects would have resulted in increased washback behaviors in the classroom. However, her responses on the questionnaire conflicted with a low frequency of washback behaviors from the Spanish 104 tests in the classroom.

4.8 Cathy

4.8.1 Observations

Cathy’s observations overall were characterized by her extensive past teaching experience, and her personal preference for grammar instruction. During the Winter
quarter, changes were observed in her instruction that reflected her increased knowledge of Spanish 104 and the Spanish 104 test. In the Autumn quarter, Cathy’s behaviors focused extensively on explicit grammar instruction with the students. She frequently used activities that were recycled from teaching an elementary level course, and it was common to find entire class sessions devoted to direct grammar instruction.

Cathy had a lot of experience teaching, not just in the Spanish department at the Ohio State University, but in other teaching environments at different levels. She believed that all grammar was important to students. Although the tests in Spanish 104 didn’t focus extensively on grammar, Cathy seemed to teach as if grammar was the most considerable part of the test. Her behaviors included developing extensive individual and group activities to practice the grammatical concepts. Some of these were oral activities, some written, and others a combination of the two. In the Autumn quarter it was common to go to an observation of Cathy’s class expecting to observe a test impacting lesson based on a textbook reading, and instead find a lesson focused entirely on grammar. For one class Cathy emailed the grammar explanation sheet to the entire class, and asked them to print it and bring it to class. During the class, she then proceeded to put the information from the sheet, which was in English, on an overhead transparency for the class session. Cathy then proceeded to present the grammar and discuss it in Spanish, prior to beginning the practice activities that accompanied the information sheet.

Cathy also used techniques and materials from previous courses. One behavior that was observed was her use of white boards. White boards are dry erase boards that students can write on with a dry erase marker. The boards can be cleaned off with a paper towel. In her class, Cathy was observed using whiteboards exclusively for
grammatical purposes. She had previously used these in the elementary level classes, and found them useful for gauging student comprehension of grammar concepts. She told me that she liked to use the white boards “to check for understanding (Madeline Hunter), and then adjust instruction based on that”. During the Autumn quarter Cathy did not use white boards for other purposes besides grammatical instruction and practice. In Cathy’s case, this observed behavior seemed to stem more from Cathy’s personal preference than anything else. She acknowledged during interviews that she was a strong proponent of grammar. Based on her interview remarks, it did not appear as though Cathy’s attention to grammar was as strongly related to Spanish 103 test washback as it might have been for other teachers, perhaps because of her extensive experience teaching. Though the tests at the Spanish 104 level allotted an average of thirty to thirty-five percent to grammar, Cathy continued to devote a large portion of instructional time to this area, making it an important focus of her Spanish 104 class.

During Winter quarter Cathy’s behaviors emphasizing grammar during classroom instruction lessened. She did not eliminate grammar instruction; however, it was less common to see an entire class session focused on de-contextualized grammar. More often in this quarter, Cathy used the grammar activities from the text, which were already contextualized within the theme of the chapter. In her classroom, these activities took the form of oral practice activities, written activities, and fill-in-the-blank activities linked to specific grammar pages in the text. She spent less time using the specific textbook grammar pages (Páginas Azules as they are referred to), and more time using the textbook activities that allowed practice of the oral skills. On various occasions, Cathy had students work in groups of 3 or 4, or work with a partner to complete the activities
included in the text. In this quarter, Cathy was the only teacher observed who used the
*Diario de Actividades* during a class session, and this was to check some homework
orally. The entire time spent was 13 minutes.

Although during the Winter quarter, Cathy still spent considerable instructional
time on grammatical accuracy, she was able to compensate somewhat by including more
reading and writing activities within the context of the course. Her instructional
behaviors reflected washback from the Spanish 104 tests, as direct test-like activities
were included as a part of her instruction.

As a result of student concerns about the tests, and Cathy’s increased knowledge
of the testing expectations, she spent more instructional time on the textbook readings in
the Winter quarter. Cathy, like Anne, incorporated increased time for teaching reading
strategies to students. She focused on the importance of reading the title, interpreting the
meaning, on vocabulary, and on story organization. Also like Anne, Cathy stressed the
importance of the *Pequeño Diccionario*, the small glossary provided prior to the reading
defining Spanish words in Spanish, and containing simple illustrations. This glossary
also provided the meaning of the word within the context of the story. By focusing on
reading strategies, Cathy’s students were more prepared to read, and appeared more
skilled at reading. Therefore, her students seemed more apt to use these tools when
reading. They also seemed more prepared to discuss the readings during the class and
complete class activities that focused on the readings.

Cathy shifted her instruction in the Winter to spend more time developing
students’ skills. Students began to view reading as an integral part of the intermediate
classroom, and they began to realize that grammar was no longer the primary focus of the
curriculum at this level. Although Cathy still focused on grammar during many sessions, she incorporated more direct test-like activities that reflected reading and writing during this quarter.

During the Winter quarter, discussion and interpretation of the textbook readings became an important part of the behaviors that focused on reading. For example, Cathy was effective at being able to incorporate her own questions along with the Preguntas de Orientación in order to further probe the students’ understanding, and their ability to communicate this understanding. She was also able to relate the pre-reading activities to the students’ context, and therefore able to successfully scaffold the readings for the students.

It was also more common to find Cathy asking students to write summaries of the stories they had read, and to break into groups and have each group be responsible for summarizing and explaining their part of the story. Students also practiced writing and and class sessions included time spent talking about how to write well in Spanish. Students shared their writing with the class either on the chalkboard or on an overhead transparency, and with the teacher. Additionally, students in Cathy’s class engaged in reading activities that included pre-, during-, and post-reading activities, and literary analysis of the readings from the chapter. In the Autumn quarter, Cathy exhibited few washback behaviors that focused on these activities, but in the Winter quarter these were quite apparent.

Reading and writing were an integral component of the Spanish 104 classroom curriculum. The study teachers demonstrated that they understood this by the time the Winter quarter observations were concluded; however, the amount of increased
instructional time on these skills between quarters varied for each teacher. Observations revealed that Cathy spent considerably less time than Anne, Diane, and Elena on skill development, mostly because she still granted significant time to grammar instruction. Although the percentage of her instructional time devoted to reading and writing represented an increase from the Autumn quarter, the increase was less than that of the other teachers.

4.8.2 Interviews

Cathy’s statements during interviews confirmed some of what was seen in her classes. Contradictions to observations that were revealed in the interviews provided further insight into the reasons behind some of her instructional decisions. Especially during the Autumn quarter, Cathy revealed that her students had different perspectives with respect to class instruction than what she had.

As mentioned in the observation section, Cathy spent considerable time on grammar instruction. When asked at the beginning of the Autumn quarter about her understanding of the nature of Spanish 104 and her plans for the course, Cathy stated:

… my teaching approach is the communicative approach in general, and by that I mean I try to encourage the students to speak in Spanish as much as possible…in terms of lesson preparation, I think I’m back to square one, I have to start from zero. That doesn’t mean that I can’t use materials or activities, or plans that I’ve used before, but for me, it’s going back to square one and really approaching the class with a fresh outlook and new ideas.

When Cathy was asked during the Autumn quarter how she planned to prepare her students for the tests, she gave the following response:

I plan to prepare my students for the test by directing them to the website for the, not only for 104, but also the website on, the things that you can click on the calendar where they pull up self-tests. In addition to that, I plan on announcing
the test format for them, since for these students it’ll be the first time that they’ve had this type of format, so I will explain what the test will consist of, first a reading section where you’ll have to do this, then a grammar section where you have to do this. Ah, third, I am going to prepare the students for the test by doing as many practice examples on like a review day as I can, usually in a fun way such as a game or a competition or something that they’re reviewing, or maybe re-learning something that they didn’t maybe remember the first time, in a worry free, or almost worry free environment.

Based on her comments, it would appear that Cathy had a solid understanding of the testing expectations, and she seemed confident in her approach to preparing her students for the testing program. However, later in the very same interview she said:

…I would like to see some copies of previous tests, for the purpose of familiarizing myself with the format and being able to familiarize my students with the format because in order for, in order for a test to work, and I think that tests can work or they can not work, I think the students have to be prepared, but I think the teacher has to be prepared as well. …I need to really put myself in the position of the test makers and say what is the primary objective of this particular part of the test?...And then ask myself after that, am I accomplishing that in class? Is, are part of my class activities what I, what the objectives on the test are trying to seek out.

Cathy appeared to understand how she needed to plan her instructional time; however, as revealed in the observations, she devoted more instructional time to activities that did not constitute the main objectives of the testing program. Cathy first realized a difference between her perception of class instruction and her students’ after she questioned her students about their feelings toward the instruction in the class. She gave her students a questionnaire that offered them the opportunity to express concerns, and tell her if they felt needed more reinforcement in certain areas of the class to better prepare for the tests. In her final interview of the Autumn quarter, she revealed that she had been surprised to discover that students believed that their needs were not being met:

…several students made the comment that they thought we needed to spend more class time on the reading, and more review of the readings before the test since
that was what the test was primarily about. And I thought I had spent a decent amount of time on the readings, but apparently several students didn’t think it was enough. …I did change my plan at a certain point to devote more time to the readings, to devote more time to preparing for the readings on the test, doing activities in small groups, not just going over the questions, but doing things like, Ok, this group come up with a summary, a ten sentence summary about the reading. This group come up with ah three or four questions about the reading. So I think that I tried to do more for the readings, because they were such a major part of the tests. Not to say that I ignored the grammar, because of course that’s what I think, what, that’s a very big part of the class. And that’s so important as well.

Observations during the Winter quarter confirmed what Cathy said in the interview. She did include more test-like activities, such as summaries, and student generated questions; however, she did not ignore grammar. Both of these aspects were readily observed in her classes.

As mentioned in the section describing Cathy’s observations, her extensive teaching experience, her well developed teaching repertoire, and her personal preferences influenced her instruction more than anything else. Cathy referred to Madeleine Hunter at some point in every interview, as she commented on how her teaching had been influenced by the ideas of this educational researcher: “I practice modeling. I really believe in Madeleine Hunter’s instructional theory into practice. I use a lot of that, a lot of checking for understanding.” Her affinity for Hunter’s ideas was evident in her classes as she incorporated activities designed to check for students’ understanding. Cathy’s comments on one of her checking for understanding activities are as follows:

…I have these little 2x2 dry erase whiteboards, and I give each student one of those. We get paper towels and a marker, and most of that is used to do grammar, but that was a perfect, perfect, perfect way for me to assess what they know. I would ask …a grammatical question “¿qué te gusta hacer?”[what do you like to do] reviewing “gustar”, [to like]and they would just write down their response, and boy, when everybody holds it up, you know exactly who knows this concept and who doesn’t. So therefore, if I see a lot of wrong answers from really good
students, and everybody else, I’m going to know I better stop and explain that response.

Again, Cathy revealed that her primary focus for activities like this revolved around grammar. Because the researcher noted that Cathy spent considerable time on grammar instruction, she questioned her about how she chose class activities, and what constituted typical class activities. Cathy responded with the following comments:

… activities are instruction and then practice where usually, with regard to grammar, I will explain a grammatical concept, after we’ve looked at it contextually, and then they will practice that context, or that construct, whatever. …when we do our review, I don’t devote specific time just to vocabulary, it’s more on grammar and the readings. I don’t want to say that I have less attention to things that are not as represented on the test, however if I see a test for example, and it’s sixty percent reading, then I’m going to need to spend a lot of class time on doing those readings. And I think that sometimes that’s difficult because I’m coming from teaching in .01 and .66 classes where really the most things you do are grammar. So it’s trying to, I’m personally dealing with trying to not do so much grammar, I mean you observed the class today, it was almost entirely grammar and we were supposed to do reading today.

The reason that Cathy’s behaviors seem to be heavily influenced by her past teaching experience and personal preferences emerge in the following comments. Cathy emphasized that grammar was an aspect of language learning that she particularly enjoyed, even though she recognized that she needed to balance out her interests with the demands placed on the students by the Spanish 104 course and its testing program. This may explain the washback behaviors from the Spanish 104 test that Cathy seemed to exhibit during the Winter quarter. In the wake of student concerns, and her own increased knowledge of the Spanish 104 testing program, she accepted that she needed to include direct test like activities as part of her instruction.

…I would say that I’m more aware of what it, what, what weights the test has in certain areas over others. And even though I would love, because I love grammar, I’d love to spend all the time doing grammar, those readings are really
important because they’re the majority of the test, and really this class is on listening and reading, and that’s what I need to do most of.

Although some of Cathy’s changed behaviors in the Winter quarter appeared to reflect Spanish 104 test washback, other behaviors seemed to reflect increased knowledge and understanding of the Spanish 104 course. The distinction between these behaviors was difficult to distinguish. Speaking about the transition between the elementary level and the intermediate level, Cathy commented:

[from] …the instructor standpoint, I think the change for me was not really so bad, and I don’t know if that’s the result of just me, or the fact that I’ve been teaching for nine years, which might be part of it. I think that maybe a good way to expose instructors to what they’ll be doing in 104…is to share what you have. …So I think that maybe more activities towards what you actually have to teach in 104 and how to come up with these cute little games, or classroom activities would be a good idea for students, and I know that we had the 104 workshop, and I know we did a couple of things, but I remember it was more about explaining to us what 104 was, and how the book worked, and how the layout, and that’s very important, but I would have liked to have more. …I also think that people that come into this program….having never taught before, and I think that there’s so many things you have to remember, that I think, you know lesson planning sometimes get pushed off to the side until the very, very last minute. …maybe a little bit more explanation on how to lesson plan…

It was apparent that Cathy was referring to lesson planning within the context of the material for Spanish 104, as she perceived the content of Spanish 104 demanding different skills and abilities from teachers and students than previous courses have demanded. Because Cathy’s experience extended beyond the scope of the elementary and intermediate programs, it does not seem to suggest that her changed instructional behaviors in the Winter quarter were simply the result of emerging teacher behaviors and a developing repertoire. Rather, they also exhibit washback from the Spanish 104 testing program. As Cathy learned more about the Spanish 104 testing program, and more about the students’ expectations with regard to that testing program, she focused more on the
test in her instruction. This is supported in comments that Cathy made during the Winter quarter:

I spent a lot more preparation time thinking about what activities I could do as far as the reading, since that’s the majority of the test. And I tried to implement those in class as much as I could. …For example, we would break into small groups and have each group summarize a particular paragraph, or get the main idea of a particular paragraph. And then we would write it out together up on the projector thing [Elmo®], or on the computer so that everybody could see that, and I think that was the one thing that was the main difference that I did in this class because I wanted to do more practice activities in preparation for that portion of the test.

It has been acknowledged that Cathy believed that grammar was important, and she devoted significant class time to grammar instruction. However, her awareness of the need to develop and focus on the skills of reading and writing for test preparation was apparent. Her comments made during the final Autumn and Winter quarter interviews were confirmed during Winter quarter observations as she balanced grammar instruction with more emphasis on reading and writing. She commented on the shift:

I mean we usually spend, the majority of the time I spend is dedicated to, I’d say the biggest chunk is dedicated to preparing them for the reading because I think it’s the most difficult aspect of the test, because they have to remember the reading whereas they don’t have it in front of them. And also with the grammar, I think it’s sometimes a little bit easier on the test to do grammar because often times there’s an example.

Cathy further remarked,

I think that I was guided more by an awareness of what the students are, need to do for the test, and I wanted to communicate that awareness to them. … …I wanted them to be aware of what the test was like, how it was set up, and I’m not saying that this dictated my whole class because it did not, but always the activities that we did, for reading especially, were really geared towards the test. …I think that in the Autumn quarter I just kind of said, well, we went over the test topics and I said this is what’s going to be on the tests, and of course we had a review day, but the review day was um equally divided in terms of grammar, reading, reading passages, and the practice reading that they get. Whereas in the Winter quarter, I think I took a lot more time to explain what the test was like, and what they should expect. And something else that I did…and I did this a little bit
in the fall, but I did this a lot more in the winter, was I would say to the students, ‘Ok, how can you best study for the reading portion of the test? Give me some ideas’, and then people would give me some ideas and then we would practice some of those in class.

4.8.3 Cathy’s Questionnaire Results

For Cathy’s questionnaire, her average mean response was 3.77, a response rate that is above neutral and slightly below agree for most questions. The questions with the responses that revealed the most agreement were questions four and seven. The results for Cathy’s mean responses to the questionnaire are in Figure 4.3.

![Cathy's Questionnaire Results](image)

**Figure 4.3.** Total mean responses on questionnaire for Cathy

With respect to learning strategies that Cathy would recommend to her students, she responded to each prompt with a four or a five (agree; strongly agree). Cathy indicated that she felt all of the strategies were important, and especially agreed with the emphasis on the skills of speaking and reading.

For question seven, which asked about how to motivate the students to learn Spanish, Cathy again responded most strongly with four and five (agree; strongly agree),
indicating strong agreement for using more interesting language games, giving students more encouragement to learn, and providing students with effective language learning strategies. She also commented that she felt that changing the schedule of the classes to meet only three or fours days a week would help students to be more engaged. Her students frequently complained to her about the fact that Spanish language classes met daily.

The most influential factors on Cathy’s teaching, prompts rated with a five, were professional training, teaching experience and beliefs, past experience as a language learner, the need to obtain satisfaction in teaching, and slightly less important, but still rated with a four (agree) were exams and learner expectations. The importance Cathy gave to exams and learner expectations was apparent in her classroom during Winter quarter observations, and in her comments during interviews. Cathy accepted the importance of tests within the context of the student experience. Although she commented on more than one occasion that she did not want to teach to the test, she conceded that her students had a very real expectation that they needed to be informed and prepared for tests, and they viewed this as one of Cathy’s responsibilities to the class. In the Winter quarter, Cathy spent 51% of her observed instructional time focusing on textbook reading and writing activities, and 11% of her observed instructional time on general skill building activities; both of which help the students prepare for the tests. Her responses on the questionnaire supported what was seen in the classroom observations. She demonstrated more washback behaviors during the Winter quarter.
4.9 Diane

4.9.1 Observations

Diane’s observations were characterized by an eagerness to maximize the use of Spanish 104 materials and prepare for the tests, along with influence from her past teaching experience in the Autumn quarter. Winter quarter observations revealed a heavy reliance on the textbook, and an increased focus on grammar instruction.

During the Autumn quarter, Diane’s instruction suggested that that she was attempting to learn as much as possible about the Spanish 104 course and the tests. Her instructional behaviors indicated that she appeared to be trying to familiarize herself with the textbook materials, while at the same time intending to incorporate her own materials to the context as well.

Diane often used numerous transparencies in her classes. These transparencies would sometimes be activities copied from the De pasco text, and other times activities that Diane developed. During the Autumn quarter, she spent significant time on skills and made great efforts to incorporate all aspects of the textbook chapter components, and test components in her classes. In these classes, activities devoted exclusively to listening comprehension, vocabulary in context, grammar in context, and textbook readings were observed. Diane appeared committed to maximizing the course materials available to her, and making sure that her students were prepared for the tests.

In contrast to the other teachers, Diane’s students focused on skill development by using the pre-reading activities, the during reading activities. When time permitted, Diane also included post-reading activities that were modeled after text activities, but original to her. She often made literary analysis a consistent and focused part of these
classes. The students would work in small groups to analyze themes from the textbook readings, to write summaries of the selection, and/or to write about the meaning of the text. At times Diane would collect this written work from the students, read it over, and make comments and suggestions for corrections and improvement, and then return the work to the students. Sometimes Diane would take excerpts from the student written summaries and write them on a transparency. Then she would spend 10-15 minutes of class time going over the grammar or the excerpt, or discussing suggestions to improve the written expression. Diane appeared focused on students’ skill building by including frequent behaviors focused in this area. These skills were needed for the Spanish 104 tests.

During the Winter quarter Diane continued to have students write summaries of the stories they had read, and break into groups and summarize a part of the story. Students also acted out some of the stories that they had read in small groups, and in front of the class. Students in her class still engaged in pre-, during-, and post-reading activities from the text, and literary analysis of the textbook readings. Her instructional behaviors placed continued emphasis on reading; however, she appeared to rely more on the activities from the text, and less on activities that she developed herself. During the Winter quarter she appeared less adept at manipulating the given activities to her individual class. She appeared to use textbook activities more in the exact form in which they were presented. She still used textbook reading and writing activities that continued to provide tasks focused on the skills of reading and writing.

Although during the Autumn quarter the teachers in the study focused a lot on grammar, Diane’s teaching behaviors showed fewer signs of grammar focused instruction
during observations, and she spent less of her total instructional time on this topic than others (22%). Her instructional behaviors indicated a smooth transition between the elementary and intermediate levels, specifically Spanish 103 and Spanish 104. Most of her grammar instruction observed during the Autumn was contextualized within the theme of the chapter, or the specific lesson for the day, and instruction was frequently linked directly to one of the textbook readings. Diane appeared to work hard at making sure that grammar instruction was not isolated.

Diane’s behaviors changed significantly during the Winter quarter, and not as expected. She increased her focus on grammar. Her grammar instruction during this quarter incorporated many grammar activities that she used in prior teaching that did not relate to the theme of the chapter, or the readings in the chapter. For example, although she used grammar practice activities from the textbook like Beth, she also employed pair work activities (A/B information gap activities), activities from Spanish 103 courses that she had instructed, and overhead transparencies contextualized within themes of a different textbook altogether. Although Diane did use more grammar activities from the De paseo text than she did in the Autumn quarter, she also frequently supplemented textbook activities with worksheets or overheads that she had used in other teaching contexts. There were many instances in Diane’s Winter quarter classes when it was obvious that a worksheet had been taken from a textbook that was not part of the Spanish program at Ohio State. The results of these behaviors were class sessions packed with grammar activities on multiple topics. The transitions between the activities were rapid, and often abrupt, leaving students confused and sometimes unsure of what they were supposed to be doing, or why they were doing it. Diane’s behaviors indicated that
grammar was a predominant focus for instruction in her Spanish 104 class during the Winter quarter, as the percentage of instructional time that she devoted to these activities increased.

The classroom observations exposed the unanticipated changes that occurred between quarters in Diane’s classes. She began the Autumn quarter more focused on skills, and exhibiting more washback behaviors from Spanish 104 tests than other teachers. But during the Winter quarter she moved away from including more test-like activities, toward including more direct grammar instruction, and more de-contextualized grammar practice. Diane’s comments during her interviews show that perhaps she overestimated her students’ abilities, while underestimating her responsibilities. This may be associated with why she reverted back to the more familiar teaching practices from Spanish 103, and revealed less behaviors indicative of washback in the Winter quarter.

4.9.2 Interviews

Diane’s interviews provided information that conflicted with some of the classroom observations, but at other times provided possible explanations for her unexpected change in behaviors between Autumn and Winter. The first signs that Diane may have overestimated her students’ ability became apparent in one of her Autumn quarter interviews:

....I think I expected almost too much in terms of what they would be able to do in 104 at first, and I didn’t think I’d have to put down my ah, such structured exercises- take from this column, take these verbs. You know, and I still need to do more of that, ah, I can see, because they’re not really, uh, I guess they didn’t do enough of that in 103.66 overall.

Diane was enthusiastic about teaching Spanish 104, saying, “I think this course is more interesting because they’re asked to apply it a lot more, it’s not just fill in the
blanks…And I think people have to put more into it”. She added though that she believed that students needed to have a better sense of the course content and expectations prior to enrolling in the course in order to

…let them know, especially in 103 and 103.66, what 104 is going to be like before they take it. And I think that that might help them. Maybe even to let them know at all levels, when you get to 104 you’re going to have to do the same thing. Just to give them a little bit of a warning that they should be aware of what it’s going to be like.

She further commented, “…I feel like I had that expectation to slap up a pretty open ended question and let them go with it, but they weren’t able to do that”.

When Diane talked about her plan for teaching Spanish 104 during the Autumn quarter, she mentioned exactly what was observed in her class sessions. She talked about trying to incorporate as many activities as possible during class sessions, and she gave careful consideration to developing students’ skills, especially in the area of listening, reading and writing; the skills that constitute the majority of the exams. She specifically stated in terms of grammar:

…then, of course, with the grammar you know, I’m trying not to put up, you know, the verb charts, but I have them ready, you know, and I can move them up there. But I tell them, ‘they’re in your book, you don’t need to copy this’. But I start right out really incorporating with questions using the grammar first, and then if we have to do the aside, very quickly do the aside, but you know, I’m not up there to teach grammar and to show the verb charts. I don’t want them to see grammar as separate. I think that with the earlier levels it seems so very disjointed that at this point we need to incorporate it.

Based on these attitudes, the shift toward increased grammar during the Winter quarter seems even more surprising. But in her final interview during the Autumn quarter, Diane discussed things that she wanted to focus on for the Winter quarter. When she was
questioned about how her teaching changed after Test 1, she replied:

The main thing that I changed was the literary analysis in dealing more with the stories, and my focus shifted entirely in that direction. But now, I don’t know. What I really want to work on over break I think, is getting more of the pair activities with the grammar incorporated and embedded more. Ah, I don’t feel like I did that very well with the grammar, so…

Although Diane’s comments revealed that she believed she needed to focus more on grammar, she did not necessarily indicate the reason why. Based on comments she made during her Winter quarter interview, it appears that the reasons were likely linked to the students’ performance in classroom activities, and their ability to write on the tests. Regardless of the reason, in her Winter quarter observations Diane focused more on grammar activities.

Diane’s comment to the question about changes made in her teaching during the Winter quarter since Test 1 of the Autumn quarter directly referenced grammar. “Ok, the major thing I tried to do was I wanted to have more pair activities that incorporated the vocabulary but most importantly the grammatical structures in some kind of context”.

When she was later asked about specific test-like activities that she included in her classes during Winter quarter, she answered,

…I tried to do a lot more as I said with pair activities with grammatical structures, because I found that they had trouble with them. It seemed like fall quarter, I felt that they were ready for the grammatical portions of the test, whereupon I wasn’t satisfied with the scoring on the grammatical portion.

Although it was not the main focus of the tests, Diane’s perception was that grammar was the area of the test where students demonstrated the greatest weakness. The students’ weak skills appeared to manifest themselves in the their writing, which was a significant
part of the chapter tests. When talking about how she helped the students prepare for the writing on the tests, Diane said,

…I try to get error correction built in where I have transparencies of error correction or else I have them rewrite for error correction. Because there are some students that just make the same mistakes over and over and over again in writing. And even though I’m trying to tell them, when you have a lot of students, it’s hard. Particular students who want to see a change, and I’m still grasping a little bit with that kind of thing, they, they’re not real certain, some of them still make the basic errors that I’m trying to shake out.

Further acknowledging the weaknesses of her students, she said,

I enjoy teaching 104, however, you know the expectation is that they can say more things and that they can participate in a manner much higher than 103.66 or 103.01 because realistically they just don’t have the tools.

Although Diane stated that she wanted to include more contextualized grammar activities in her Winter class sessions, observations showed that grammar became direct and often unfocused. Diane spent so much time on grammar concepts that she seemed to overwhelm the students during many of the class sessions. Although she had demonstrated many behaviors suggestive of Spanish 104 washback during the Autumn quarter, her behaviors during the Winter quarter revealed less of a tendency toward washback. She incorporated fewer direct test-like activities, and more activities based on what she viewed as students’ weaknesses. Her classes incorporated many more practices common to the Spanish 103 level class, and her behaviors signaled a return to teaching methods more appropriate for that curriculum.

4.9.3 Diane’s Questionnaire Results

For Diane’s questionnaire, her average mean response was 3.0, a response rate that is neutral for most questions. It must be noted that Diane did not follow the instructions for completing the questionnaire because she did not rate each of the
statements. Instead she rated the prompts in order of one to five, and left the other prompts blank. Therefore, because more than half of the prompts were not rated, her mean response rate was affected. As a result of her rating system, each of her questions yielded a completely neutral response; however, some of her written comments lent support to the behaviors observed in her Winter quarter classroom teaching, and what she revealed in her interviews. Additionally, the fact that some prompts were not rated at all by Diane indicated the lack of importance she gave to them in terms of influencing her teaching and her beliefs. The mean questionnaire response for each question for Diane is reported in Figure 4.4.

![Figure 4.4. Total mean responses on questionnaire for Diane](image)

Diane indicated that the most significant aspect of extra work placed on her teaching in Spanish 104 was trying to organize more focused activities that reflected exam activities. During Autumn quarter classroom observations, she was seen spending significant time trying to focus on the required readings and the skills. But she increased
the time she spent on grammar instruction markedly in the Winter quarter, a somewhat
contradictory action to her belief that the most significant aspect of extra work was
related to creating exam like activities. Although grammar was a test activity, as
previously indicated, it was not the major focus of the tests. A possible reason for this
discrepancy may be explained by her response to question three.

Diane’s most significant responses about the most challenging aspect of teaching
Spanish 104 were: the students’ current low level of Spanish, inadequate time for
students to practice Spanish outside of the classroom, and students’ lack of reading and
vocabulary skills. For her, these presented the greatest challenges to teaching Spanish
104. Recalling that Diane also made frequent references to student ability in her
interviews, perhaps this explains her focused emphasis on grammar instruction during the
Winter quarter- she believed that the students needed to learn the grammatical structures
in order to be successful in the language.

The factors that Diane indicated as most influencing her teaching were: teaching
experience and beliefs, and the need to obtain satisfaction in teaching. She did not rate
exams, the syllabus, learners’ expectations, peers, and/or supervisors as affecting her
teaching. Because Diane had significant teaching experience prior to teaching in the
Department of Spanish and Portuguese at Ohio State, her response to this question may
have differed considerably from those of less experienced teachers. During the Autumn
quarter, she appeared confident in her ability to navigate a new syllabus, and make sense
of a new testing program because she was an established teacher. However, her
confidence seemed to diminish during the Winter quarter when her instruction showed
more signs of behaviors typically associated with Spanish 103, and her class activities reflected the Spanish 104 test less.

4.10 Elena

4.10.1 Observations

Elena’s class observations were characterized by her past teaching experience, and her commitment to meeting students’ expectations. Because Elena had already taught courses focused on reading and writing before, she seemed to find teaching Spanish 104 easier than some other teachers. She regularly incorporated activities that focused on reading and writing in her class. Elena did this by using the textbook activities and her own activities. She would often stimulate class discussions on the textbook readings that lasted an entire class session.

Elena was effective at incorporating her own questions along with the Preguntas de Orientación in order to further probe the students’ understanding of the readings, and their ability to communicate their understanding. She was also able to effectively relate the pre-reading activities to the students’ context, therefore successfully scaffolding the readings for the students.

More than other teachers, Elena exhibited frequent behaviors focused on textbook readings by continually finding ways to bring up the characters from the stories, or incorporate details from the stories in the daily class activities. The discussions and activities that she included in her classes offered her students multiple opportunities to ask questions, to demonstrate understanding, and to practice their reading and writing skills.
Elena’s behaviors incorporated focused time for teaching reading strategies to students beginning in the Autumn quarter. Like Anne and Cathy, she provided instruction that focused on the importance of reading the title, and the information that the title alone could provide for the reader prior to initiating reading an unknown selection. A vivid example of this was during a test impacting lesson focused on an excerpt about Gloria Estefan, titled *Nuestra Gloria*. Elena focused class discussion on possible interpretations of the title within the context of what the students had read. She helped the students to go beyond the literal link to the name, and led them to see how, within the Latino culture, people are proud of Gloria Estefan and her contributions to the Latino population. Elena wanted her students to see the meaning and the impact of “Our” Gloria.

As did other teachers, Elena stressed the importance of the resources in the textbook, such as *Pequeño Diccionario*. This resource has been mentioned in other teacher’s observations as the small glossary provided prior to the reading defining Spanish words in Spanish, with simple illustrations. To reiterate, this glossary provided the meaning of the word within the context of the story. Because Elena focused on these strategies, her students appeared to make better use of the textbook resources, and they appeared to be able to read in the target language with less anxiety than students in some of the other classes.

In both Autumn and Winter quarters, like other teachers, Elena asked her students to write summaries of the stories they had read, and to break into groups and have each group be responsible for summarizing and explaining a part of the story. Her students also practiced writing, and class sessions focused on how to write well in Spanish.
Students in her classes also regularly engaged in reading activities that included pre-, during-, and post-reading activities, and literary analysis of the textbook readings. The main difference between quarters with respect to these activities was the time frame. In the Autumn quarter Elena sometimes focused her activities in specific areas of reading and writing after students demonstrated a need for assistance, or an actual weakness. During the Winter quarter, Elena was able to anticipate potential areas of weakness, and she was therefore more proactive with her instructional behaviors. She was confident that her students needed specific skills to be able to meet the demands of the tests, and she began focusing on these skills sooner. Her washback behaviors from the Spanish 104 tests reflected an emphasis on skill development.

Elena’s students both quarters viewed reading as an integral part of the intermediate classroom, and they realized that grammar was not the primary focus of the curriculum at this level. During class sessions, Elena’s students were observed engaging in conversations about the stories with the teacher, and with the other students in the class. They appeared to ask fewer clarification questions about the stories, and about subsequent activities that focused on post-reading activities.

Elena also spent considerable instructional time consciously linking reading to writing. Observations of Elena showed that she spent more class time than other teachers developing students’ skills in the areas of reading and writing within the context of the required readings. During the Autumn and Winter quarters, Elena’s students engaged in activities in and out of class (homework) that encouraged the development of these skills. As mentioned earlier, a common test activity was to ask students to write a summary of one of the stories they read. During both quarters this was a typical activity in Elena’s
classes. For one lesson she also chose to focus time on how to write a coherent narrative, and a summary in Spanish by modeling them. She wrote out a personal narrative on the chalkboard, explained it to the students, and had the students work in pairs to develop and complete their own narrative. Students also put examples of summaries on the chalkboard or the overhead machine. Students in Elena’s class were introduced to useful writing techniques, and phrases and vocabulary that served to facilitate writing in Spanish. Elena exhibited noticeable behaviors in the area of skill building that were indicative of washback from Spanish 104 tests, given that these tests focused largely on skills.

Many of the other class observations revealed that teachers spent a large percentage of time focusing on grammar instruction, especially during the Autumn quarter. Elena did exhibit behaviors focused on grammar, but it was more common to find her grammar instruction embedded in lessons focused on reading or writing. She spent less total instructional time on grammar than the other teachers. Her behaviors reflected the shift in emphasis between the Spanish 103 and 104 tests, and she appeared to make the transition with more ease than others.

During the Winter quarter, Elena still included grammar instruction in her classes. Additionally she borrowed some of the grammar reference sheets that other teachers had shared with her, and she distributed them to her students. This behavior seemed to be the result of two things; one, her students’ desire to have information about the grammar presented in a manner easy for them to understand, and two, what she learned from the other teachers about how native speakers of English sometimes interpret Spanish grammar. Remembering that Elena was a native speaker of Spanish, she reported it was
challenging for her to comprehend why students made certain mistakes. Error patterns that may be attributed to interference from English as a native language were not always obvious to Elena. When native English speaking teachers shared their grammar reference or explanation sheets with Elena, she thought they were worthwhile to share with her students. It was her intention to help her students understand the Spanish language better. Although on the surface her attention to these types of grammar worksheets may seem to reflect behaviors typical of Spanish 103, for Elena it was more a case of wanting to address the students’ needs. Her students informed her that they wanted information about the tests, and about grammar, and she saw these reference sheets as a way to give her students what they wanted. Even though she used the review and reference sheets, Elena still believed that increased use of skills would result in improved grammar abilities. During an interview she talked about how the tricks she learned from the “American professors” were useful for some of her students.

4.10.2 Interviews

During her interviews, Elena provided information that confirmed what was observed in her classes. She was focused on developing the students’ skills in Spanish, and she was committed to meeting the students’ needs in Spanish 104.

Unlike the other teachers in the study, Elena was a native speaker of Spanish, now living in the target language country. Therefore, although she might have experienced some of the same challenges that many of the students faced while learning a foreign language, her language experience was more intense than most students experience at the Spanish 104 level. But as a result of her experiences, Elena indicated that she offered learning strategies from her own language experiences to her students to try and facilitate
their learning. She wanted to meet the students’ needs. Elena saw reading as a tool to help her own language learning, and communicated this to her students. Given the focus on reading in Spanish 104, and the difficulty that students sometimes experienced when having to read more in Spanish, being able to relate to reading challenges helped Elena reach her students. She said,

…I [had] big problems myself when I was learning English. And I remember having to figure it out by myself because my teachers, the only thing they could tell me was ‘you have a problem’. ‘Well, I know that, how do I fix it?’ So I had to solve that problem by myself. …So, that meant a lot of work, mainly reading because I had no contact with English directly, so dictionary work and grammar stuff was useful, especially when I found short texts on the same subject matter I could start identifying some patterns carrying certain meaning.

Elena viewed skill development as a means to build grammar accuracy, based on her own language learning. She shared her own language learning with her students to help them make connections with what they were doing in Spanish 104. When Elena was asked whether she believed that students should engage in more reading and writing in the elementary program she added:

…if we did more reading and those types of activities to kind of build the vocabulary and build those skills in the lower levels, it would help them [students] have a smoother transition into the intermediate [level]? Yes, that’s at least what…were the ways in which I could figure out my own problems. And then when I was teaching English I was [sic] try to discover which was, because I was just doing it, and the main advance in the language having no contact with native speakers, was when I had to study for my career as a translator, history and literature. I had to read huge novels, and I had to use the dictionary. And I used both a translation dictionary and …meaning dictionary.

Elena was the only teacher in the study who overtly demonstrated behaviors reflecting skill development in both her classroom instruction and her interviews during both quarters from the beginning.
Elena indicated that she believed it was important to keep the students’ needs in mind when planning for instruction. When she first began teaching Spanish 104 in the Autumn quarter, she sought out experienced 104 teachers to help her better understand the demands of the course.

…it’s my first time in 104. …I have heard of some activities other classmates have done, ah…I talk to DR and she told me about activities she develops in classes, some of the people like D that has, she has more time than me…Now I’m just learning, this first course I think, is just like going to be learning the material, getting organized, and also, grasping the time thing, that sometimes I think I can do, I don’t know teach four different activities and then I can not.

Along with the demands that the course placed on the teacher, Elena was also concerned about the demands the course placed on the students. She made several references to these demands, acknowledging the importance that students placed on the tests. Her first comment is about the transition between Spanish 103 and 104:

…I think that it could be a little bit difficult for them…that it is as you said, ah, they have much more reading, much, much, much more reading and reading take[s] a lot of time, and sometimes they, if they took the other courses, then they can feel that this is too much for them.

She further added,

…I know that they have to pass the exams, and they are going to be more interested if they pass exams, so ah, that is a need they have. And, my, my, what I want is them to learn my culture. Try to transmit them something meaningful, that is not their main interest, it’s mine. So I need to find a middle point between the two things, so when I, when they see that I am interested in their needs, then they just, their barriers come down and say, ‘ok, maybe she has something interesting to tell us. Maybe it’s not that bad.’

Passing tests was a priority to her students, and Elena intended to provide instruction that helped her students meet this goal during both Autumn and Winter quarter.

One of the changes observed in Elena’s classes was the time frame in which she introduced activities focused on skills in the Winter quarter. She consciously focused on
reading and writing in the Autumn quarter, but typically after students demonstrated weakness on a test, or difficulty during classroom activities. She said that she

focus[ed] their attention in linking words and …what I tried was mainly…to go on focusing those things because I [had] already done [them] from the very beginning after the first exam. So, the …the other two exams just confirmed me where the problems, where the problem was…

During the Winter quarter, Elena reported that she took less time to introduce activities focused on these skills, and observations affirmed what she said. She was able to focus and attend to students’ needs in a more timely way than before because she was more alert to what to watch for. She stated,

…I guess that, well the Autumn quarter was my first quarter, and then during the second time I had to teach this I started to focus on the things I had to cover, [the things that were]…difficult for the kids during the first one. So for example, all of the stuff about Sak-Nicté, how to make sentences and how to use connectors, and that material, I use[d] it before they had to write summaries. So I include[d] it as part of the teaching, not afterwards as correcting what they have [already] done.

Elena anticipated the test demands and her students’ skill, and included activities that reflected these demands and skill levels. Elena exhibited Spanish 104 washback behaviors to a greater extent than other teachers in the study.

It is possible that Elena’s behaviors may be a combination of increased knowledge of the tests, as well as the course and the curriculum. She reported that she felt more comfortable teaching the class for the second time during the Winter quarter. However, she includes in her comment, the importance of the tests to the class; therefore, it is unlikely that all of her behaviors are the result of increased familiarity with the course.
Well one thing is that I feel sure what I am saying now. So I can, I noticed that me being sure about things makes them feel better and then that I can handle the material with more flexibility than at the beginning. At the beginning I was kind of following day after day, the material, the schedule. I notice that I know the material, I know how the exam goes, then I know that I can manipulate it. It doesn’t matter this or not, what matters is this. I know the nucleus, the core, so I can manipulate it.

She also said, “maybe, ah, the second time I knew exactly what direction the exam went. So I kind of tried to focus on those things.” She had a better understanding of the tests and the expectations, and she felt confident that her class instruction prepared students to meet these expectations. This is what Elena wanted, and she believed that this was also what her students wanted:

...I cannot impose my interest over their interest. Their interest is to pass the exam. So my strategy is always to show them, ‘ok, I am interested that you pass the exam, I am here with you for that. And once I do that that, they say, ‘Ok that is cool, she’s interested in me, why not be interested in her.

4.10.3 Elena’s Questionnaire Results

For Elena’s questionnaire, her average mean response was 4.10, a response rate that is slightly above agree for most questions, and the highest overall agreement rate of all of the teachers. Questions five and seven yielded the highest responses for her. The mean response for each question for Elena is illustrated in Figure 4.5.
For question five Elena indicated that she considered the most effective types of activities for language learning: language games, role-play and group discussion, authentic materials, and extracurricular activities. She also considered training in basic language knowledge important, while she remained neutral on task oriented activities. Class observations showed signs of using activities focused on basic language knowledge, such as the attention Elena gave to writing, and the functions of specific words and phrases.

For question seven, focusing on ways to motivate students to learn Spanish, Elena’s only neutral responses were using more authentic materials, and giving students more encouragement to learn. All of her other responses were rated strongly agree, and they were using more: mock exam like activities, real life language activities, more interesting language games, positive attitude toward language learning, and providing students with effective language learning strategies. Judging from her responses to this
question, it seemed as though Elena believed that balancing out what she considered to be important and central to successful language learning, with what students wanted, was what kept students interested. Her interview comments further supported this perspective. The behaviors that Elena demonstrated in her class sessions supported her responses on the questionnaire.

The factors that Elena reported as most influencing her teaching were: teaching experience and beliefs, past experience as a language learner, the need to obtain satisfaction in teaching, exams, learners’ expectations, supervisors’ expectations, and social expectations. Her response showed her acceptance that students have specific needs and expectations in the language classroom, and a crucial part of her belief system was acknowledging, and working towards meeting these needs. Elena admitted that students’ needs largely reflected tests, and therefore she spent significant class time answering students’ questions about the testing process, and on preparing students to meet the demands of the tests. In her classroom Elena demonstrated Spanish 104 washback behaviors that were a result of not only her knowledge of the testing program and its focus on skills, but also of the students’ desire to know as much as possible about the tests and the expectations.

4.11 Summary of Observations, Interviews, Teacher Questionnaires

The observations, interviews and the teacher questionnaire provided information about the behavior patterns teachers demonstrated in their classrooms. The patterns signaled a variety of factors influencing teacher behaviors, including how tests influenced the instruction and the class activities in Spanish 104. The following sections provide
summary information that details the patterns and trends identified in the observations, interviews, and the questionnaire, in addition to further information about teacher beliefs.

4.11.1 Summary of Behavior Patterns in Teacher Observations

Analysis of the teacher observations exposed patterns that influenced the instructional behaviors seen in the Spanish 104 classroom. These patterns reflected instructional behaviors focused on grammar, behaviors focused on textbook reading and writing activities, behaviors focused on reading and writing skills in general, and behaviors focused on past teaching experiences coupled with washback from past testing experiences. Behaviors that could be directly linked to tests in Spanish 104 were designated as demonstrating washback. One of the guiding questions in determining these behaviors was posed by Watanabe (2004). This question asked: if the tests did not exist, would the behaviors that teachers exhibited be the same?

4.11.2 Teacher Behaviors Exhibited Focusing on Grammar: Autumn and Winter Quarters

Classroom observations revealed that teacher behaviors that focused on explicit grammar for the most part were due to the familiarity that these teachers had with teaching this topic, and in some cases the result of washback from the Spanish 103 tests. It is difficult to distinguish between the two, because washback manifests itself differently for different teachers, and in some cases, behaviors appear to have been a combination of teachers’ familiarity with teaching grammar in previous courses, and the tests in previous courses.

Each teacher in the Department of Spanish and Portuguese, via a series of workshops (801 workshop for new TAs, and the course Spanish 801: Teaching Spanish at the college level), learns various techniques and activities for incorporating different
types of grammar activities into classroom instruction. For all of the teachers in the study, their training for incorporating grammar and grammar activities into an elementary communicative class was more extensive than their training for teaching reading and writing in an intermediate communicative class. Therefore, it is not surprising that observations indicated that most instructors devoted significant class instruction to the textbook grammar topics. These topics included grammar that students had already learned in Spanish 103, or earlier classes.

The teacher behavior that focused on specific grammar instruction mostly included the following topics: the use of preterite and imperfect tenses, the use of subjunctive tense, the use of indirect and direct object pronouns, the use of the conditional and the imperfect subjunctive tenses, and comparisons of equality, superiority, and inferiority. All of these grammar topics had previously been studied by students in the elementary levels, and the students had access to extensive practice activities in their textbooks, in their workbooks *Diario de actividades* for *De paseo*, and on the supplemental website for the text ([http://depaseo.heinle.com](http://depaseo.heinle.com)).

The classroom activities in Spanish 104 that accompanied this instruction included tasks such as working with partners on paired A/B activities, situational activities such as practicing a structure by completing a task (*Dear Abbey*), working in groups to complete exercises and then writing answers on the chalkboard to discuss as a class, having mini quizzes that focused on the grammar points, and/or completing worksheets, or grammar exercises from the textbook. None of these activities were direct test activities. Teachers justified the focus on grammar by acknowledging the fact that the grammar included in the textbook chapters was a part of the test. During the Autumn
quarter especially, teachers also seemed to believe that students struggled with many of the grammar topics. The students participated easily in these grammar activities. This was most likely due to the fact that they had extensive experience with these types of activities at lower levels, specifically Spanish 103, in the Spanish language program at the Ohio State University.

Table 4.2 quantifies the grammar related behaviors that teachers exhibited in the Autumn and Winter quarter. The total number of minutes from the eight classroom observations that each teacher devoted to grammar activities, in addition to the percentage of total class time that was devoted to this topic of instruction are presented. As the teacher observations detailed, Anne, Cathy, and Elena reduced their grammar instruction in the Winter quarter, while Diane increased her grammar instruction. Beth showed no change in the amount of instructional time she spent on grammar between the two quarters. The complete details of the instructional distribution for each class observation are in Appendix D.

<table>
<thead>
<tr>
<th>Instructional Behavior: Grammar Focused</th>
<th>Total minutes of class time from 8 class sessions Autumn 2004</th>
<th>Total percentage of class time for 8 sessions Autumn 2004</th>
<th>Total minutes of class time from 8 class sessions Winter 2005</th>
<th>Total percentage of class time for 8 sessions Winter 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>216/384</td>
<td>56%</td>
<td>91/384</td>
<td>24%</td>
</tr>
<tr>
<td>Beth</td>
<td>176/384</td>
<td>46%</td>
<td>176/384</td>
<td>46%</td>
</tr>
<tr>
<td>Cathy</td>
<td>258/384</td>
<td>67%</td>
<td>133/384</td>
<td>35%</td>
</tr>
<tr>
<td>Diane</td>
<td>86/384</td>
<td>22%</td>
<td>118/384</td>
<td>31%</td>
</tr>
<tr>
<td>Elena</td>
<td>91/384</td>
<td>24%</td>
<td>55/384</td>
<td>14%</td>
</tr>
</tbody>
</table>

Table 4.2 Total minutes and percentage of total class time from eight class sessions devoted to grammar instruction during Autumn and Winter quarters.
As table 4.2 illustrates, a large portion of instructional time was devoted to teaching grammar during the Autumn quarter in three of the five classes. This would also suggest that for some teachers, washback from previous Spanish 103 tests played a role. The focus of the Spanish 104 test, and the objectives of Spanish 104 were not centered on grammar. Grammar, however, was a significant part of the Spanish 103 course and its tests. It is logical that as teachers planned their classes, they focused on activities that mimicked these tests. It is plausible that the grammar behaviors during the Autumn quarter were related to teacher’s familiarity with teaching direct grammar, either in the elementary program in the Department of Spanish and Portuguese, or in other teaching experiences. This study did not intend to determine the line of demarcation between these two factors; however, it is inaccurate to say definitively that behaviors were solely the result of one or the other.

During the Winter 2005 quarter, the teacher behaviors that focused on grammar tended to shift in two distinct ways. They either became a less dominant focus of the classroom instruction, and/or they became more contextualized within the theme of the chapter and the chapter readings. More often in Winter quarter, teachers used the grammar activities from the text, which are contextualized within the theme of the chapter. These activities consisted of oral practice activities, written activities, and fill-in-the-blank activities linked to specific grammar pages in the text. Overall, less instructional time was spent on using only the specific grammar pages (Páginas azules as previously described), and more time was spent on using activities that fostered practice of the oral skills. In many classes, the teachers had students work in groups of 3 or 4, or work with a partner to complete the activities from the text. One example comes from
instruction focused on Chapter 5 from the textbook. The theme of this chapter, *El Medio Ambiente*, focuses on the environment. In several classes, students discussed ways that cities suffer from pollution, talked about ways that they would clean up the cities (employing the conditional tense), or posed recommendations for the future generations to protect the environment using the imperfect subjunctive, (ie., grammar points from the chapter). Although students worked with grammar, they talked about the theme of the chapter, and typically one or more of the readings from the chapter. It is presumptuous to classify these behaviors as pure washback only from Spanish 104 tests, because we can assume that teachers will make better use of the resources available as they become more familiar and more comfortable with the course. However, the students were focused on the skills of reading, writing, and speaking within the context of the chapter theme, two of which are direct skills needed for the test. The activities and the instruction complementing these behaviors were related to test content, and therefore are considered as helping to prepare the students for the tests.

During the Winter quarter there was less use of the *Diario de actividades* during class time. This workbook is designed to give students practice with the skills necessary for understanding the information from the text. It provides extensive practice in listening comprehension, writing, and the grammar structures contained in each chapter. It was common during Autumn quarter to observe teacher behaviors where the grammar activities from the *Diario de actividades* were used in class correcting homework, practicing the structures in groups, and writing answers on the chalkboard or the overhead. Teachers in these cases demonstrated behaviors that seemed to result from their familiarity with using similar activities in Spanish 103. Teachers may have also
believed that the grammar and structure activities from the textbook would model the
grammar and structure activities from the tests, something that is common in the
elementary level classes.

During Winter quarter, work from the *Diario de actividades* was presented as
independent work that students needed to undertake to practice their skills. All of the
teachers still assigned activities from the workbook; however, their expectation was that
students would complete this work on their own, outside of class.

4.11.3 Washback Behaviors Focused on Textbook Reading and Writing Activities:

**Autumn and Winter Quarters**

Teacher behaviors displaying instruction centered specifically on reading and
writing activities from the textbook appeared to indicate washback from the Spanish 104
tests. This is because the tests focus heavily on textbook readings (30-40%), and there is
a significant amount of writing that students must complete on the tests. Therefore, using
these activities extensively directly reflects test expectations, and helps to prepare the
students for the expectations of the tests.

The latest edition of the Spanish 104 text (*De paseo*) used in the Spanish language
program includes a section with authentic readings, and a section with activities focusing
on contextualized activities promoting reading and writing. The teacher’s edition of the
text provides extensive support materials for teachers. An example of the support
materials from the Teacher’s Edition is as follows:

New in this edition, the **Tercera Etapa** has been renamed the **Lectura** and
now contains one cultural reading (**Lectura cultural**) and one literary reading
(**Lectura literaria**). Students learn practical strategies for reading in the
**Sugerencias para la lectura** section and are guided through the reading process
with pre-, during-, and post-reading activities. The unique **Pequeño diccionario**
precedes each cultural and literary text, and each video and listening selection to help with students’ comprehension. Every literary reading is accompanied by an author biography, key vocabulary words and phrases in context, comprehension checks, and activities that encourage debate and discussion. Along with literary selection, Análisis literario helps students to develop critical reading skills (Long, Macián, 2005, p. IEA-5).

In Autumn quarter, despite access to the textbook support materials, the amount of instructional time spent on reading and writing activities was limited for most of the teachers. Overall, during the Autumn quarter the teachers did not maximize textbook activities promoting reading and writing with their students as a part of their regular classroom instruction. Therefore, they did not engage in direct test-like activities, so they did not appear to exhibit instruction marked by notable washback behaviors in this area. More often, instructional behaviors included assigning the Preguntas de Orientación, questions that accompany each reading and are found in the margin next to the specific reading. These questions are designed to “guide students through the selections and to develop their critical thinking skills (Long, Macián, 2005, p. IAE-1-)”. The most common way teachers used the Preguntas de Orientación was assigning them for homework along with the reading. The next day in class, the teacher asked the students to compare their answers in groups with other students, or one on one. Some teachers never discussed the answers to these questions; they simply asked if the students had any “questions or problems (preguntas o problemas).” Others went over the answers together with the students, sometimes using the overhead projector with the answers written on a transparency, or using the chalkboard. With the exception of Elena, the teachers used these questions directly from the textbook and did not build on them to further probe student comprehension.

150
Table 4.3 illustrates the teacher behaviors focused on reading and writing suggesting washback from the Autumn and Winter quarter. The total numbers of minutes from the eight classroom observations that each teacher devoted to textbook reading and writing activities, in addition to the percentage of total class time that was devoted to this topic of instruction, are presented.

<table>
<thead>
<tr>
<th>Instructional Behavior: Textbook reading and writing</th>
<th>Total minutes of class time from 8 class sessions</th>
<th>Total percentage of class time for 8 sessions</th>
<th>Total minutes of class time from 8 class sessions</th>
<th>Total percentage of class time for 8 sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>58/384</td>
<td>15%</td>
<td>170/384</td>
<td>44%</td>
</tr>
<tr>
<td>Beth</td>
<td>51/384</td>
<td>13%</td>
<td>69/384</td>
<td>18%</td>
</tr>
<tr>
<td>Cathy</td>
<td>18/384</td>
<td>4%</td>
<td>197/384</td>
<td>51%</td>
</tr>
<tr>
<td>Diane</td>
<td>55/384</td>
<td>14%</td>
<td>207/384</td>
<td>54%</td>
</tr>
<tr>
<td>Elena</td>
<td>152/384</td>
<td>40%</td>
<td>175/384</td>
<td>46%</td>
</tr>
</tbody>
</table>

Table 4.3 Total minutes and percentage of total class time from eight class sessions devoted to instruction of required readings included on tests during the Autumn and Winter quarter.

Diane and Elena used some of the pre-, during-, and post-reading activities exercises in the book. For the other teachers, it was more common to move on to a grammar or structure item, or if was the end of the class, to mention the homework that usually included one of these topics. The teachers appeared to exhibit few washback behaviors in this area in spite of the focus on the textbook readings, and the expectation that students could not only comprehend and interpret what they read, but also articulate the information effectively in Spanish.

By Winter quarter, the teacher behaviors emphasizing test-like reading and writing activities from the textbook were more apparent in all classes. Anne and Cathy,
especially showed dramatic changes. This increased emphasis on reading and writing activities indicates focused washback from the tests. Because the skills of reading and writing accounted for 51% of the test, both teachers and students engaged in behaviors that focused more on skills needed for the tests during the Winter quarter.

All teachers were more familiar with the testing process than they were in the Autumn. Based on student reactions from the Autumn quarter, this familiarity helped them to anticipate what students needed to know and to practice in order to be prepared for the tests. Therefore, instead of just asking students to complete the *Preguntas de Orientación*, teachers more frequently asked students to expand upon these questions. For example, students completed activities such as “what would the characters say”, they wrote summaries of the stories that they read, they acted out the stories that they read in front of the class or in small groups, and they wrote their own questions that were comprised of different types of questions (factual, inference, personal opinion, and reaction). All teachers indicated that for the exam, students needed to be able to comprehend and answer questions, and interpret the stories. However, only some teachers included more comprehension activities for the students prior to the exam.

Although not all teachers were confident about the best way to prepare their students for tests, using these activities encouraged students to develop their reading and writing skills. With the exception of Beth, the teachers no longer gave the impression that they were satisfied students understood the story based solely upon the answers to the *Preguntas de Orientación*.

Discussion and interpretation of the reading became an important part of the behaviors that focused on reading. Being able to effectively and accurately discuss and
interpreted (in writing) the textbook readings was a direct test activity. Therefore, teachers’ behaviors focused in this area were considered to be washback from the Spanish 104 tests. All teachers in the study increased their instructional time on these activities, with Cathy demonstrating the most significant change.

4.11.4 Teacher Behaviors and Possible Washback Focused on General Skill Building In Reading and Writing: Autumn and Winter Quarters

It was difficult to determine how much of teacher behaviors focused on general skill building were washback from Spanish 104 tests, and how much resulted from teachers being better prepared to teach skills after having taught Spanish 104 the previous quarter. For Anne and Cathy, their behaviors indicated more washback because the teachers articulated a clear understanding of the importance of skills for the tests, and skill development as a tool for language learning. Yet for Beth and Diane, their behaviors focused on skill building seemed to be more of an increased familiarity with the course and less of a direct result of the test. This is because these teachers admitted that they still did not feel adequately skilled to prepare the students for the tests.

Regardless of the amount of instructional change due to washback, and the amount due to course familiarity, the teachers’ changed behaviors demonstrated recognition of the link between skills and the tests. The familiarity with the course allowed the teachers teach the Spanish 104 curriculum more effectively, and to better prepare the students for the tests.

Table 4.4 illustrates the total numbers of minutes from the eight classroom observations that each teacher devoted to developing general skill building in reading and writing during the Autumn quarter. The percentage of total class time that was devoted
to this area of instruction is also presented. Lessons that incorporated and expanded upon
the activities from textbook post reading activities, as well as teacher generated activities,
were categorized as general skill developing activities. The complete details of the
instructional distribution for each observation session are in Appendix D.

<table>
<thead>
<tr>
<th>Instructional Behavior: General Skill Building</th>
<th>Total minutes of class time from 8 class sessions Autumn 2004</th>
<th>Total percentage of class time for 8 sessions Autumn 2004</th>
<th>Total minutes of class time from 8 class sessions Winter 2005</th>
<th>Total percentage of class time for 8 sessions Winter 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>27/384</td>
<td>7%</td>
<td>94/384</td>
<td>25%</td>
</tr>
<tr>
<td>Beth</td>
<td>46/384</td>
<td>12%</td>
<td>94/384</td>
<td>25%</td>
</tr>
<tr>
<td>Cathy</td>
<td>15/384</td>
<td>4%</td>
<td>41/384</td>
<td>11%</td>
</tr>
<tr>
<td>Diane</td>
<td>62/384</td>
<td>16%</td>
<td>37/384</td>
<td>10%</td>
</tr>
<tr>
<td>Elena</td>
<td>33/384</td>
<td>9%</td>
<td>106/384</td>
<td>28%</td>
</tr>
</tbody>
</table>

Table 4.4 Total minutes and percentage of total class time from eight class sessions devoted to general skill building in the areas of reading and writing during the Autumn and Winter quarter.

Even though during the Autumn quarter each of the teachers devoted some time
to developing reading and writing skills, as Table 4.4 indicates, the percentage of total
instructional time devoted to this area was just 16% at its highest. Given that 66% of the
test focuses directly on skills, this revealed a sharp contrast.

Students in the classes where the least instructional time was devoted to reading
and writing skills were uncomfortable working on activities that demanded the use of
these skills. When asked to read or write spontaneously, students asked frequent
questions of the instructor after an assignment or task had been explained. Or they asked
the researcher questions such as “how would I say…?” “are we supposed to do this?...”,
indicating that they had little practice with skills.
During the Winter quarter, Anne, Beth, Cathy, and Elena increased their use of text and teacher generated pre-, during- and post-reading activities. Developing the students’ general skills in reading and writing, and not increasing their knowledge about the textbook readings, became a focus in their classes. An example from Anne, Cathy, and Elena’s classes was the way that these teachers taught the textbook readings. For many of the students the required readings that are taken from literature, such as *El guerrero y Sak-Nicté, Presagios,* and *Sala de Espera* are challenging to comprehend, to discuss, and to write about. Yet these are all activities that the students must demonstrate on the tests. Anne, Cathy, and Elena created activities that helped students make sense of the stories, and then went beyond the stories. The activities helped the students prepare for the sections of the test that demanded skills. Beth did include more activities focused on skill development; however, she appeared to struggle with how to best implement these activities in class.

Reading and writing are an integral component of the Spanish 104 curriculum. Anne, Beth, Cathy, and Elena demonstrated that they understood this by the time the Autumn quarter observations concluded. On the contrary, Diane spent considerably less time than the others on skill development, still granting significant time to grammar instruction. It is difficult to determine whether the increases shown by Anne, Cathy, and Elena were because of familiarity with the course content, or because of increased knowledge of the Spanish 104 test. Based on observations alone, it is difficult to determine the greater influence. However, when considering the additional information provided by interviews and the questionnaire, the underlying reasons that may have contributed to the changed behaviors seem more apparent. For Beth the change was most
likely influenced by her increased familiarity with the course. For Anne, Cathy, and
Elena it was likely the influence of both factors. For Elena her behaviors were most likely
a result of washback than familiarity with course content. Another factor that may have
also influenced the instructional behaviors of the teachers is the student expectations with
regard to the test.

4.11.5 Teacher Behaviors Resulting from Past Teaching Experience and Testing
Experience: Autumn and Winter Quarters

The final pattern of teacher behaviors in the classes appears to be the result of past
teaching experience, combined with past testing experiences. This complex behavior
included instruction resulting from teachers’ experiences in previous courses and testing
programs. The reason that the testing program seemed to constitute a part of the behavior
is that not all teachers had significant teaching experience necessary to have highly
developed teaching skills. With a limited past teaching experience, it is unlikely that the
driving force behind instruction consisted solely of past teaching experiences. Rather, it
would also reflect past testing experience.

Past teaching experience behaviors were demonstrated most often during
instruction related to grammatical topics, and the use of past teaching materials. As all
teachers focused on grammar during the Autumn quarter, some quite heavily, the specific
amount of time focused on specialized grammar points appeared to be reflections of what
each individual teacher deemed to be the most important grammar concepts, as well as
those that were included as parts of the testing program either at the Spanish 103 or
Spanish 104 levels. To reiterate, the specific grammar points that received this focus
included the appropriate use of two past tenses in Spanish, and the use of the subjunctive
mood (which produces a significant change in Spanish). These concepts are difficult for native speakers of English to grasp and apply effectively. This is in part because of the nature of the difference between the two languages, and because of the fact that native speakers of English studying Spanish at this intermediate level commonly make comparisons between the foreign language and their native tongue. Because these tenses are taught and tested extensively at the Spanish 103 level, the teachers in this study had experience teaching and testing these concepts. As a result of the teaching and testing experiences, the present Spanish 104 teachers made comprehensive review sheets detailing uses and rules of these grammatical concepts, practice worksheets, and overhead transparencies with explanations and practice activities. These sheets were modeled after previously taught material in other levels, and did not consistently reflect the context of Spanish 104.

The use of comprehensive review sheets signaled past teaching behaviors. These review sheets were described in the observations section as those sheets that included an acronym or mnemonic in order to help students remember the grammar information presented. Several of the classes used these sheets that were from when the teacher had taught Spanish 103, or Spanish 102. Teachers believed that students in Spanish 104 struggled with the same grammatical topics and concepts. Consequently, they spent significant time presenting information and techniques to students that they believed would help them to master the grammar. Because each of these teachers had most recently taught in the elementary program where the curriculum focused more on grammar, these behaviors reflected teacher’s past teaching experience at that level.
Another behavior reflecting past teaching experiences was planning instruction with topics from past courses. For example, during the Autumn quarter Anne, Beth, and Cathy frequently used overhead transparencies, worksheets, or handouts that were from a Spanish 102 or Spanish 103 course that they had taught. The activities on the overhead transparencies or worksheets consisted of student work in pairs, or information gap activities (activities divided into roles [A/B] where student A must provide information for student B, and vice versa). Frequently it was possible to discern the chapter in the Amistades text (Spanish 101-103) that these activities came from, as a result of the nature of the grammar topic and the researcher’s familiarity with this text. For example, although the theme of Chapter 2 in De paseo is El Yucatan (the Yucatan), and one of the grammar topics is the use of the preterite and the imperfect tenses, Beth used the activity El Oso (The Bear) to practice the tenses. This activity is from the context of the Amistades textbook, and has no relation to the theme of the Yucatan.

The relation of activities and/or instruction to other courses and tests was also apparent due to the fact that some activities used in the Spanish 104 classroom were not related to, or contextualized within the chapter theme of, the De paseo text. Although teacher behaviors reflecting past experience were most often related to grammar, it was the use of materials contextualized to previous courses instead of the present course, that set these behaviors apart from a focus on grammar activities taken from the Spanish 104 De paseo text.

In the Winter quarter there were was much less demonstration of teacher behaviors stemming from past teaching, and past testing practices. Behaviors observed during the Winter quarter were more balanced with activities that reflected both what the
present Spanish 104 curriculum, and the Spanish 104 tests demanded of students. Behaviors focused on explicit grammar instruction were still apparent, and there were some class sessions that still focused predominantly or entirely on grammar instruction and practice. These classes however, used grammar practice activities from the Spanish 104 text (*De paseo*) more often. This contextualized the grammar for the students within the framework of the chapter. These behaviors indicated a shift from the Autumn quarter when it was common to see behaviors that focused on grammar activities and instruction from previous texts.

The washback behaviors in the Autumn quarter from past test experience seemed diminished in the Winter quarter. Behaviors now appeared to reflect the present test. These behaviors included the use of explicit test activities such as discussion of and activities with new readings, listening comprehension activities, and focused writing on test topics.

A trend that may have been related to the decrease in possible teacher behaviors from past teaching and past testing practices was that teachers followed the course calendar more closely during the Winter quarter. As previously described, the calendar is developed to follow the course content in a logical and coherent manner, in order to provide opportunities for students to develop and enhance their skills. It appeared that after teaching the course once, the teachers acknowledged that the calendar was a helpful tool. It was much more common during the Winter quarter than the Autumn quarter to observe the lesson described on the calendar. This in turn made it much easier for the researcher to determine the amount of time spent on specific activities. For example, the researcher chose to observe the lesson focused on the reading *Nuestra Gloria* because it
was a test impacting lesson. In the Autumn quarter, observations included class sessions focused on oral vocabulary activities, oral grammar activities, written grammar activities, and one discussion and presentation of the actual reading. The activities used were contextualized within the theme of the chapter, *La Música*. In the Winter quarter, 4 out of 5 teachers (Anne, Cathy, Diane, and Elena) focused completely on the reading, and used various activities throughout the entire lesson directly as specified on the calendar. Perhaps teachers realized that following the calendar and including the activities from it could help them better prepare students for the demands of the tests. The researcher witnessed that as teachers began to adapt their lessons and instruction more closely to the calendar, they also exhibited more behaviors indicative of washback from the present test. They used direct test activities frequently in class, and admitted to giving more time to test preparation. The teachers also appeared to be more organized, more confident in their abilities, and better at manipulating the materials in the text. For example, they found it possible to incorporate thematic grammar activities after a discussion of the chapter readings because they were able to build upon the activities completed in class. After discussing a story, a teacher might ask the students to write 5 sentences about “what would you do?” in this situation, or “what would the characters say in the following situations?” The students used the grammar from the chapter, but they were also writing about the story that they had read and discussed. Therefore, in addition to focusing on the grammar, students also focused on further developing their interpretation and understanding of the story. They practiced their written expression, and it was something they needed explicitly for the tests.
It is difficult to decipher the quantity of teacher behavior that was a result of past teacher training and experience, and that which was a result of past testing experiences. What seems apparent is that teachers relied heavily on past teaching experience when they taught Spanish 104 for the first time. Also, the patterns and behaviors that influenced much of the instruction during the first time teaching largely resembled patterns and behaviors that reflected demands of testing programs from different levels, specifically the Spanish 103 level.

4.11.6 Overview of Interviews: Autumn and Winter quarters

Several common themes emerged from the teacher interviews. Some of the details revealed by teachers supported what the researcher observed, while other information presented itself as a complete contradiction. The first interview began by asking each teacher about her proposed plan for the Spanish 104 course, and how she envisioned the course progressing. This was followed in subsequent interviews with questions that sought to determine the reality of what was actually happening in the course. These interviews took into account the teacher’s exposure to the testing program and the course materials, and the teacher’s interaction with the students compared to the teacher’s stated hopes and intentions during the first interview.

4.11.7 Themes: Autumn Quarter

The common themes from the Autumn quarter interviews were like many of themes that had manifested themselves in the observations. They included the teachers’ past teaching experiences, the teachers’ perspective on grammar in language learning, the teachers’ familiarity with course content and curriculum and how this affected her preparedness to instruct Spanish 104, and the teachers’ perception of the students’
preparedness for Spanish 104. The major difference between the interviews and the observations was that sometimes what a teacher shared in interviews was not what she displayed in her class.

4.11.8 Teachers’ Past Teaching Experience

All of the teachers interviewed made references to the other courses that they had taught prior to Spanish 104, and stated that they believed Spanish 104 would be significantly different from previous courses in terms of the activities used in those courses and appropriate activities for 104. Most added that their approach to teaching the course was “communicative”, meaning that they intended to have the students use the target language as much as possible during class sessions.

The emphasis on the communicative approach for teaching probably came, in part from training in the Department of Spanish in Portuguese. The Department’s philosophy is that all language classes should be taught in the target language. Teachers are taught to use the target language as the medium for instruction, and all training and instructional feedback emphasizes this. These teachers all had a minimum of three quarters teaching in the Ohio State program so the concept of using the target language for instruction was a part of their teaching repertoire.

In the first interview, teachers showed a tendency to declare independence from the elementary level courses preceding Spanish 104. In the second interview, this attitude was replaced with a perceived need to utilize materials from those previous courses. After the first test was administered, teachers were faced with the reality of what students were expected to know for the exams, and what students actually were
learning through instruction. In the second interview, teachers reported using more of the activities and materials that they had used in previous courses.

Thus, the past experiences of these teachers influenced the behaviors that dictated how teachers planned instruction in their Spanish 104 classes. For those teachers who had the bulk of their teaching experience in the Spanish department at Ohio State (Anne, and Beth), their instruction reflected the most recent previous courses that they had taught (Spanish 102 and Spanish 103/103.66). Even though Spanish 104 was a different course, and the tests are noticeably different than other courses, teacher behaviors modeled courses and tests that the teachers found most familiar. The teachers with additional teaching experience from another context (Cathy, Diane, and Elena), also showed signs of their previous experiences as they planned for instruction in Spanish 104. For all teachers, interviews and classroom observations confirmed that past teaching experience had a significant impact on teacher behaviors.

4.11.9 The Role of Grammar in Language Learning

Four out of five of the teachers in this study (Anne, Beth, Cathy, and Diane) revealed that they believed grammar was an integral, if perhaps not the most important component of language learning. And for each, a significant focus on grammar instruction was apparent in the classroom. Interviews revealed that teachers believed that grammar had to be a major component of instruction, and Cathy indicated a personal preference for grammar, which influenced her attention to grammar in the classroom.

4.11.10 Instructor Preparedness and Familiarity with Content and Curriculum

All of the teachers acknowledged that they felt they did not have sufficient training to familiarize them with the Spanish 104 curriculum and its testing demands
prior to teaching the course. Each teacher had attended an afternoon training workshop held at the end of the Spring quarter before their first time teaching in the Autumn quarter. The Academic Program Specialist who was the Supervisor for Spanish 104 and the Spanish Language Program Director conducted the workshop. During this workshop, all received the textbook and materials for the course, samples of how the text and supplementary materials functioned, the expectations for the course, and an overview of a test in Spanish 104. The participants were also provided with information about the Spanish 104 website which provides information such as the course calendar, the course requirements, the grading policy, and additional information for both students and teachers.

Throughout the Autumn quarter, instructors were not provided with additional support unless they requested it from other teachers, or from the Spanish 104 supervisor. Accordingly, as the course progressed throughout the Autumn quarter, the teachers became more aware of their level of preparedness to teach Spanish 104, and their ability to prepare students for the tests in 104.

By the end of the Autumn quarter, each teacher admitted the impact her preparedness had on her teaching. Teachers remarked how their preparedness influenced their methods for preparing students for tests, and how they anticipated their teaching would be modified for the second time teaching the course in the Winter quarter. Additional teacher comments from all teachers during the Winter quarter teacher interview that are detailed in the next section, revealed even more about the overall teaching experience between the two quarters and the changes that teachers self reported.
4.11.11 Student Preparedness

All teachers reported that student preparedness for the course influenced their behaviors and instructional practices in the classroom. Every teacher mentioned the level of student abilities, and how her assumptions associated with student abilities affected instruction, and subsequent instructional changes made throughout the quarter. Anne, Beth, Cathy, and Diane commented that the expectations regarding student ability that they had when they first began teaching Spanish 104 were incorrect. Because teachers initially believed that students would be able to do more with the language, and with a greater degree of accuracy, they planned classroom instruction based on these assumptions. When teachers realized that their expectations were not always aligned with the students’ abilities, they appeared to compensate by reverting to behaviors more common in lower level classes. It appeared as though teachers associated a weakness in student abilities with a need to match instructional behaviors to the courses where the students supposedly acquired these abilities.

4.11.12 Winter Quarter Interview

One interview was conducted with each teacher during the Winter quarter at the end that quarter. The purpose of the interview was to gather data about the changes in teacher behaviors that occurred as a result of familiarity with the course curriculum, and the testing program as self reported by the teachers.

4.11.13 Theme: Winter Quarter

The predominant theme from the one interview conducted during the Winter quarter arose from the increased test awareness that instructors had at the beginning of the Winter quarter. This test awareness made the most significant impact on the
instruction that students received during that quarter. Winter quarter interview comments reflected an acknowledged change in focus resulting from increased familiarity with the course and the testing program for Anne, Cathy, Diane, and Elena. The interview indicated that all teachers had an increased comfort level with the course materials as a result of their familiarity with the course curriculum, and the testing program.

The Winter interview also indicated that the student expectations with regard to the test influenced the class activities and the instructional behaviors of the teachers. The class activities were affected in different ways for different teachers, as were the instructional activities. Despite the differences, all teachers did take the students’ expectations regarding tests into consideration.

4.11.14 Familiarity with Course Content and Materials Related to the Test Resulting in Shifts in Instructional Emphasis

Because teachers were more familiar with the curriculum and the content included on the tests, Anne, Cathy, Diane, and Elena reported that they significantly changed how they prepared students for the tests during the Winter quarter. The instruction and activities that they chose focused more on direct test-like activities, and tasks that would prepare students to successfully meet testing demands.

The change in instruction most reported was the emphasis given to reading and writing during the second quarter. As a result of learning about the emphasis of reading and writing on the exams, and watching students struggle in these areas during classroom activities and on the tests, the teachers admitted to intentionally focusing classroom instruction on reading and writing. The researcher also observed this in the classroom during the Winter quarter.
4.11.15 Student Expectations with Regard to Textbook Readings and the Tests

Teachers acknowledged the students’ expectations with respect to the tests and the textbook readings during the Winter quarter. All teachers spent time discussing the format of the test, and reviewing the possible types of questions that students might encounter on the textbook reading section of the tests. Teachers also made sure to include the textbook activities that focused on the readings during the class sessions. The extent to which teachers manipulated, developed, expanded on these activities varied according to the teacher. Anne, Cathy, Diane, and Elena spent significantly more time during the Winter quarter on activities that focused directly on test-like activities from the textbook readings as a result of their students’ expectations about the tests. This theme was most identified and observed in Anne, Cathy, and Elena’s classes, as they knew from first hand accounts (questionnaires, conversations, and student feedback) that students felt as though they needed more preparation with the readings to meet the demands of the tests. As a consequence, these teachers focused specifically on this area during instruction in the Winter.

4.11.16 Summary of Interviews from Autumn and Winter Quarters

The themes identified during the Autumn quarter interviews were related to patterns identified in the observations. Teachers’ strong beliefs about grammar influenced their instructional behaviors. Teachers focused intently on students’ preparedness for Spanish 104, and their own preparation for teaching Spanish 104, more than they focused on the Spanish104 testing program. Teachers’ comments indicated that they were aware of their individual assumptions relating to Spanish 104 and to the skill level of their students.
As the Autumn quarter progressed, the teachers became more attentive to the tests and test demands, although some were still choosing to model their instruction on their past practices. Teachers also appeared to be influenced by previous testing programs that they had experienced. The observations confirmed that most teachers were challenged by the transition to the intermediate from the elementary level class in terms of skill instruction, student expectations, and student abilities.

Familiarity with course materials and content of the tests manifested itself as increased teacher confidence and more test like instruction during the Winter quarter. Although aspects of classroom instruction still included instructional practices from the Autumn quarter, such as grammar instruction, and the use of supplemental materials from previous courses, and the teachers’ beliefs still influenced the direction of instruction at certain times, the familiarity that teachers now had for the course and the tests affected the instruction during the Winter quarter. In other words, teachers demonstrated washback behaviors from the present tests in Spanish 104. Every teacher, to varying degrees, reported focusing on activities that were directly included on the tests, and observations confirmed that every teacher focused on activities from the tests more in the Winter quarter than the Autumn quarter.

4.11.17 Teacher Questionnaire Overall Comparison

Individual teacher responses to the Teacher Questionnaire were presented in the sections devoted to each teacher. As described in Chapter 3, a comparison of all responses was generated. This was to see if there were similarities and differences between teacher responses, and/or to see if patterns could be detected that might confirm or contradict information from observations and interviews. The following section
presents the information from the teacher questionnaire with respect to the two highest rated responses on the questionnaire, and the questions that referred directly to tests.

4.11.18 Summary of Two Highest Rated Responses

Table 4.5 shows the mean response rate of all teachers for each question. The mean response to the entire questionnaire was 3.7, i.e., higher than neutral and slightly below agree overall. Although the mean responses do not reveal as much information as individual responses to specific questions, they are reported here to provide the history for each teacher, and to better illustrate the highest rated responses.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Mean Response Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>3.86</td>
</tr>
<tr>
<td>Beth</td>
<td>3.58</td>
</tr>
<tr>
<td>Cathy</td>
<td>3.77</td>
</tr>
<tr>
<td>Diane</td>
<td>3.00</td>
</tr>
<tr>
<td>Elena</td>
<td>4.10</td>
</tr>
</tbody>
</table>

Table 4.5 Mean questionnaire response for each teacher

The questions that received the highest level of agreement from teachers were questions four and seven. Question four, which was about the learning strategies that teachers would recommend to their students, received the highest response rating for the prompt “emphasis on skills”. Teachers indicated that they believed that skill development was an effective strategy for learning. The observations lent some support to this finding, as teachers with significant teaching experience tended to emphasize skills in their classes. All teachers appeared to recognize the importance of skills; however, the degree to which the researcher observed skills integrated and emphasized during classroom instruction varied according to the teacher and her experiences. Because the
skills were directly related to the tests, teacher behaviors emphasizing skills indicated positive Spanish 104 test washback. These behaviors came across as being influenced by teacher expertise and training.

Teacher responses to question seven, a question about the ways to motivate students to learn Spanish, varied with respect to the prompts that teachers rated strongly agree. However, all teachers strongly agreed that “creating a positive attitude toward language learning” was an important way to motivate students. All teachers also agreed that teaching experience and beliefs had a significant influence on their teaching, which supported what was observed in the classroom.

4.11.19 The Importance of Tests: Questions One, Two, Six, Seven, and Eight

Questions one, two, six, seven, and eight specifically referenced tests. Teacher responses to these questions revealed additional information about teacher washback behaviors that were observed in the classroom. For example, in question one, teachers were asked about the extra demands placed on them by the tests in Spanish 104. Beth, Cathy, Diane, and Elena rated the prompt “organizing more focused activities that reflect exam activities” a 5 (strongly agree), and Anne rated the prompt with a 4 (agree). During the test impacting observations during the Winter quarter, Anne, Cathy, and Elena were the three teachers who seemed to exhibit the most washback behaviors with class activities that directly reflected exam activities. Although Beth and Diane rated this prompt high, they both relied heavily on textbook activities, and did not seem to include as many direct test like activities in class. Their response implies that they perceived the pressures of the tests; however, they appeared to have been unsure how to effectively incorporate test-like activities into the class. Although the possibility exists that they
were using more test like activities during non-observation lessons, on observation days directly preceding an exam, the researcher did not observe Beth and Diane using any test-like activities.

Question two asked about the most significant changes that teachers made in their teaching as a result of the tests. Beth, Cathy, and Diane rated the prompt “teaching according to the test format” with a 4 (agree), Anne rated this prompt a 2 (disagree), and Elena rated it a 1 (strongly disagree). What is unusual about this response pattern is that it does not coincide with the researcher’s observations, or the information that came forth in the teacher interviews.

Cathy, especially during the Winter quarter, spent more instructional time teaching to the format of the test, and preparing her students for the test. Observations confirmed that she focused on the format and the material for the test, and she spent considerable time discussing strategies for taking the test. In the test impacting lessons, especially, she spent much more time during the Winter quarter exhibiting washback behaviors than she did in the Autumn quarter.

In contradiction to their responses, Beth and Diane spent the same or less time teaching according to the test format. Beth did not include more activities that modeled the test, nor did she spend considerable time talking about the test with her students. She even commented during interviews that the changes she made to her teaching were minimal, and most of them focused on including more grammar in her instruction. Diane also admitted to using more grammar in her classes during the Winter quarter. However, Diane did spend time talking about the format of the tests, and giving students test like activities of writing summaries of textbook readings. Usually these were not class
activities but homework activities, and she therefore did not spend significant class time on test-like activities in the Winter quarter. Although Beth and Diane’s questionnaire suggested that they believed they were teaching more to the test format, this was not confirmed by interviews and observations. Rather interviews and observations revealed that these two teachers exhibited limited Spanish 104 washback behaviors, especially during the Winter quarter in comparison to the other teachers.

Anne and Elena’s responses to question two also seemed to contradict observations and information from interviews. Their apparent disagreement with the statement about “using more test like activities” may have been a result of them not wanting to admit that they were “teaching to a test”. For Elena, even during the Autumn quarter, it was common to find her exhibiting behaviors that signaled washback such as students practicing actual test activities, and students spending significant time on skill building activities. Elena’s washback behaviors were sustained, and in some cases increased during the Winter quarter. For example, she not only focused on writing skills, but she wrote out an entire sample of a model composition and worked through it with the students.

As for Anne, washback behaviors in the Autumn quarter were less frequent, but during the Winter quarter they seemed very apparent. Not only did Anne make sure that students were well informed of the format of the test, she also spent considerable time working on class activities that were test-like with the intention of helping students to prepare for the tests. In the Autumn quarter Anne seemed more concerned with trying to incorporate her techniques and activities from her past teaching experiences. During the Winter quarter, she accepted that the test and skills played a large part in the Spanish 104
course, and she incorporated many more techniques and activities that directly reflected the tests, thereby demonstrating behaviors that showed signs of increased washback on her part.

Questions six and seven were related to each other, asking teachers what they thought were the major aims of students studying Spanish, and how teachers would like to motivate their students to learn Spanish. Both questions contained a prompt that referred to tests: question six “to pass examinations”, and question seven “do more mock exam like activities”. The manner in which teachers responded to the test prompt in question six appeared to be related to their response to the prompt in question seven.

Elena, Cathy, and Beth rated the question six prompt “to pass examinations” with a 5 (strongly agree), while Diane rated this prompt with a 2 (disagree) and Anne with a 1 (strongly disagree). In question seven, teachers who rated “passing examinations” high in question six, also rated “motivating with exam like activities” high in question seven, except for Beth. Elena rated the prompt from question seven with a 5 (strongly agree), and Cathy rated it with a 4 (agree). Diane and Anne gave the prompt the same rating that they gave the test related prompt in question six; 2 (disagree) and 1 (strongly disagree) respectively. Beth was the only one who was not consistent in her response, rating question seven with a 2 (disagree).

Cathy and Elena both showed signs of washback behaviors in their classrooms, and revealed information during interviews that supported the questionnaire responses. Both of these teachers mentioned that they accepted that students wanted to know as much as possible about the test, and even strategies for successfully preparing for the test. Elena acknowledged this in the Autumn quarter, and commented that she used this
knowledge as a way to get students’ attention. Because her students saw that she was willing to help them meet their expectations concerning the tests, they in turn reciprocated their attention to her lessons. Elena’s students were engaged and interested when she wanted to teach them about her own culture.

Cathy also commented that she realized that students wanted to know as much as possible about the tests. During the Autumn quarter when she perceived that her students were spending sufficient time on the textbook readings, her students indicated to her that they needed and wanted more practice in that particular area. Therefore, although her washback behaviors during the Autumn quarter were not as obvious, Cathy compensated by increasing these behaviors during the Winter quarter. Both Cathy and Elena perceived the importance that students gave to tests, and they modified their teaching as a result. They each included mock test-like activities that focused on reading comprehension, readings from the textbook, listening comprehension activities, and writing.

Beth did not exhibit more behaviors indicating washback that reflected her questionnaire response. Although she indicated that she believed a major goal of students was “to pass examinations”, she did not appear to modify her teaching to reflect this. This may explain her low rating of “using mock exam like activities”. Both observations and information from interviews indicated that Beth saw passing tests largely as the students’ responsibility. As the teacher, Beth believed she was obligated to present the course material and answer questions about it, making sure that the students understood which content material would be on the test. However, based on her instructional behaviors and information from her interviews, she indicated that she believed it was the students’ responsibility to prepare for and pass the tests. It was also the students’
responsibility to let Beth know if they had questions. In her final interview conducted during the study she commented,

…I’ve noticed, and it’s kind of frustrating as teacher when you ask a million times if they have questions and they just stare back at you. And I’ve told them, a couple of times, I said if you don’t tell me you have a question, if you don’t even say I don’t understand, I just keep going. I don’t know what else to do if you’re not helping me out here.

This comment gives the impression that although Beth knew that students placed importance on the tests, she did not necessarily perceive it to be her main goal to address this importance during class instruction. It also reflects her level of teaching experience, and her undeveloped teaching repertoire. Beth wrote on her questionnaire next to this prompt: “I don’t like the exam activities anyway; I don’t think they help the students learn.”

As for Diane and Anne, who disagreed with the prompts in both questions six and seven, the response seemed to contradict what observations and interviews revealed. Both of these teachers’ behaviors in the classroom indicated that they accepted the importance that students gave to tests, and worked to address this importance in their instruction.

Anne, while not spending much time in the Autumn talking about the tests and doing test like activities, made a conscious adjustment to her teaching in the Winter quarter. She even mentioned to me that during the Winter quarter, when she passed out a test, she went over every section with the students to make sure that they understood the instructions. She also encouraged them to ask her questions if necessary.

Diane spent time in class during both the Autumn and the Winter quarters going over the test format with the students, and making sure that the students were familiar
with the content of the tests. Although her instruction shifted during the Winter quarter, and she exhibited less behaviors suggesting washback related to the present tests, she still addressed the students’ concerns regarding the test. Diane continued to use the activities that were test like such as summary writing, and she continued to use the post reading activities from the textbook in order to prepare the students for the tests, although to a lesser degree than in the Autumn quarter. The washback behaviors that Anne and Diane exhibited were not what one would expect based on their rating of the test prompts in questions six and seven.

Question eight asked teachers to rate the factors that most influenced their teaching. The prompt “exams” was rated 5 (strongly agree) by Cathy and Elena, 4 by Beth (agree). Anne rated the prompt 3 (neutral), and Diane didn’t rate the prompt at all. For Cathy and Elena the response is logical based on their classroom behaviors and the information revealed in their interviews. For Beth, again it seems to contradict the observations and the information from her interviews. Although Beth believed that tests influenced her teaching, her washback behaviors in the classroom were limited. The response for Anne also seemed contradictory, especially in light of the increased washback behaviors observed during the Winter quarter. Although Anne indicated neutrality on the issue, her teaching behaviors suggested a fairly strong agreement with tests influencing her instruction. This influence was further affirmed in her interviews when she revealed how she changed her instruction significantly the second time around as a result of the tests.
4.11.20 Questionnaire Summary

The results of the questionnaire indicate support in some cases and contradiction in others for what was observed in the classroom, and information revealed during teacher interviews. The results provide more background about the teacher beliefs that did not always manifest themselves in the teacher behaviors. The information from the questionnaire also indicated that tests were an important influence on teacher behaviors, often in way that teachers did not fully comprehend, or seem to realize. Based on the information gained from the questionnaire, it appears that the beliefs that serve as the foundation of the teacher behaviors influencing instruction in the classroom are not always clearly understood by the teachers themselves.

4.12 Student Questionnaires

Students in the Autumn and the Winter quarter classes of the teachers in the study were asked to complete a series of questionnaires. The first set of questionnaires accompanied each chapter test. Students were asked to indicate the amount of time they studied for the specific parts of the test. The total time each student spent studying for each test was tallied for the whole class, and then analyzed and compared using Microsoft Excel®. The comparison was to investigate if students reported increased or decreased study time for certain sections of the test that corresponded to changes that teachers were observed making in their instructional behaviors related to the tests. Changes in student behaviors that corresponded to changes in teacher behaviors indicated possible student washback. The student questionnaire from the chapter tests is in Appendix C.
A more detailed questionnaire was given to the students with their final exam. The purpose of the Final Exam questionnaire was to gain information about students’ perceptions of their teacher’s behavior with regard to the test and test like activities. The questionnaire also sought to learn about the students’ perceptions of differences between the Spanish 104 and the previous Spanish classes, and students’ overall behaviors and study habits that reflected the tests.

Results from the Spanish 104 final exam student questionnaires were organized and tallied using Microsoft Excel®. Analysis focused on the questions related to the tests. Student responses from each quarter for individual teachers were also compared. The data and analyses from the students’ questionnaires provide information that suggests how students may respond to changes in teachers’ instructional behaviors related to tests, and trends or behaviors suggesting student washback. The student questionnaires also provide information for further research on washback in the intermediate Spanish language classroom. The Final Exam student questionnaire is in Appendix C.

4.12.1 Student Responses: Time Studied for each Chapter Test

After each of the three chapter tests during both quarters, students were given a short questionnaire about their study habits for that particular test. The purpose was to determine if once students and their teacher became more familiar with the testing format and the demands of the test, that students changed their study habits - in other words, if there appeared to be any washback from the testing program. Questionnaires that were not completed were discarded. Students were asked to indicate the time spent studying for each section of the test, and the overall time spent studying. Students expressed the
amount of study time in both minutes and hours. These amounts were then calculated as portions of hours, and organized using Microsoft Excel®. The following tables report the student study habits in terms of time devoted to each section for each test, for each teacher’s class during the Autumn and the Winter classes. Table 4.6 represents Anne’s students, Table 4.7 represents Beth’s students, Table 4.8 Cathy’s students, Table 4.9 represents Diane’s students, and Table 4.10 represents Elena’s students.

4.12.2 Study Habits for Anne’s Students

<table>
<thead>
<tr>
<th></th>
<th>Listening Comprehension</th>
<th>Reading Comprehension</th>
<th>Readings from Text</th>
<th>Structure/ Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AU 04</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>.96</td>
<td>1.17</td>
<td>1.44</td>
<td>.98</td>
</tr>
<tr>
<td>Test 2</td>
<td>.74</td>
<td>1.18</td>
<td>.98</td>
<td>1.03</td>
</tr>
<tr>
<td>Test 3</td>
<td>.46</td>
<td>.81</td>
<td>.93</td>
<td>.81</td>
</tr>
<tr>
<td><strong>WI 05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>.32</td>
<td>1.34</td>
<td>1.34</td>
<td>1.09</td>
</tr>
<tr>
<td>Test 2</td>
<td>.58</td>
<td>1.24</td>
<td>1.12</td>
<td>.95</td>
</tr>
<tr>
<td>Test 3</td>
<td>.44</td>
<td>1.05</td>
<td>1.00</td>
<td>.99</td>
</tr>
</tbody>
</table>

Table 4.6 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarters for Anne’s students.

Students in Anne’s class during the Winter quarter spent notably less time beginning with first test studying for the listening comprehension portion of the tests. They also spent more time studying for the reading comprehension and readings from the text portion of the test. This may be attributed to the fact that Anne’s instructional behaviors emphasized the readings more in class, and she spent more class time working on reading comprehension skills during this quarter as she had a better understanding of the demands of the test. Her teaching revealed that textbook readings and reading
comprehension were stressed much more in class instruction during the Winter quarter than they were during the Autumn quarter.

For the grammar and structure portions of the exams, Anne’s students spent similar amounts of study time both quarters. This may have been due to her behaviors emphasizing grammar in the class sessions, and/or the challenge of having to write a significant portion of the test answers in Spanish.

For the listening comprehension portion of the test, the Winter quarter students spent a third the amount of time studying for this section with the first test compared to the Autumn quarter students. But they almost doubled the amount of time spent on this section in the Winter quarter for Test 2. Anne spent much more class time focusing on listening comprehension during the Autumn quarter. Therefore, the student behaviors can be attributed to the teacher’s focus on listening comprehension between the Autumn and Winter quarters. Additionally, it may have been that after Test 1 in the Winter quarter, the students found the listening comprehension to be more challenging than they had anticipated, and therefore they decided to take advantage of the listening comprehension activities in their texts and their workbooks to better prepare for the next test. The amount of studying time devoted to listening comprehension for Test 3 was virtually the same for both quarters. This may reflect the fact that Test 3 contains information from only one chapter, and therefore students perceived it was a more manageable amount of material.

The students’ perception that Test 3 contained a more manageable amount of material may have come directly from Anne, as she commented to the students that the instructional pace would slower and more manageable. This might have reduced student
anxiety regarding test content. Another factor might also have been that student’s listening skills should have improved by Test 3, because students had been listening to their instructor speak Spanish daily during class time for approximately eight weeks. Overall, Anne appeared to exhibit washback behaviors through her use of direct test activities and skills needed for the tests. The results of her students’ questionnaires indicate that they were aware of Anne’s behaviors, and modified their study habits in response to those changes.

4.12.3 Study Habits for Beth’s Students

<table>
<thead>
<tr>
<th></th>
<th>AU 04 Listening Comprehension</th>
<th>Reading Comprehension</th>
<th>Readings from Text</th>
<th>Structure/ Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>.18</td>
<td>.74</td>
<td>.95</td>
<td>.67</td>
</tr>
<tr>
<td>Test 2</td>
<td>.87</td>
<td>1.01</td>
<td>1.36</td>
<td>1.44</td>
</tr>
<tr>
<td>Test 3</td>
<td>.41</td>
<td>.46</td>
<td>1.47</td>
<td>1.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>WI 05 Listening Comprehension</th>
<th>Reading Comprehension</th>
<th>Readings from Text</th>
<th>Structure/ Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>.37</td>
<td>.96</td>
<td>.91</td>
<td>1.45</td>
</tr>
<tr>
<td>Test 2</td>
<td>.40</td>
<td>1.20</td>
<td>1.48</td>
<td>1.64</td>
</tr>
<tr>
<td>Test 3</td>
<td>.21</td>
<td>.76</td>
<td>1.12</td>
<td>1.12</td>
</tr>
</tbody>
</table>

Table 4.7 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarter for Beth’s students.

Beth’s Winter quarter class spent less time overall studying for the listening comprehension section of the test than the Autumn quarter class. Beth was never observed teaching listening comprehension during class sessions in either the Autumn or the Winter quarter, but she did assign listening comprehension activities as homework more during the Autumn quarter than the Winter quarter. Her Autumn and Winter quarter classes reported increasing study time in this area for Test 2, perhaps because
they found the listening comprehension section on Test 1 more challenging than they expected.

Beth’s students reported changing their study habits by spending more time studying for the reading comprehension and the readings from the text sections during the Winter quarter than the Autumn quarter. This reported increase suggests possible student washback. It didn’t appear that this was due to a change in instructional behavior by the teacher, but rather a change in the background information that Beth provided to her students about the exam. Because she told them more about the expectations for the reading comprehension and the textbook readings, Beth’s students knew prior to taking the test that they would be responsible for reading a new selection and answering questions about that selection, and answering questions about the textbook readings.

Beth’s students also knew more about the types of questions that they would have to answer. Instead of simply assuming that the questions would be like the Preguntas de Orientación from the text, Beth told the students that they might have to summarize a story, or answer questions about characters from the story. Chances are that the students realized on their own the importance of the textbook readings to the test and consequently they devoted more time to preparing for these sections. Reading and the required chapter readings were the areas least emphasized during instruction in Beth’s class; therefore, students may have felt they required more preparation time. This could have contributed to the fact that the students seemed exhibit more washback behaviors than the teacher.

Grammar and structure received a marked increase in student study time from Beth’s students during the Winter quarter. This seems logical because it was an area that was stressed frequently during class sessions. Moreover, Beth held the students
responsible for grammatical accuracy to a great extent on their exams, sometimes taking off points on answers that had structural problems. Beth did this even though the test instructions simply stated that students would lose one point for each word in the answer written in English. Her students lost points on answers because the grammar of their response was inaccurate, even if it was not a stressed grammar concept from the present chapter. Given the emphasis that Beth placed on grammatical accuracy, and the amount of class time spent on grammar activities during the Winter quarter, the students’ behaviors toward increased study time is not surprising. The behavior on the part of the teacher may, in fact, have contributed to negative student washback behaviors where students became so concerned about the importance of grammatical accuracy on the tests, that they sacrificed study time in other areas. In Beth’s class, her students appeared to respond to her instructional behaviors and their own increased knowledge of the testing demands.

4.12.4 Study Habits for Cathy’s Students

<table>
<thead>
<tr>
<th></th>
<th>Listening Comprehension</th>
<th>Reading Comprehension</th>
<th>Readings from Text</th>
<th>Structure/Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AU 04</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>.75</td>
<td>1.15</td>
<td>1.26</td>
<td>1.18</td>
</tr>
<tr>
<td>Test 2</td>
<td>.34</td>
<td>1.20</td>
<td>1.35</td>
<td>.89</td>
</tr>
<tr>
<td>Test 3</td>
<td>.64</td>
<td>1.44</td>
<td>1.65</td>
<td>.98</td>
</tr>
<tr>
<td><strong>WI 05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>.55</td>
<td>1.33</td>
<td>1.27</td>
<td>1.33</td>
</tr>
<tr>
<td>Test 2</td>
<td>.31</td>
<td>1.31</td>
<td>1.59</td>
<td>1.66</td>
</tr>
<tr>
<td>Test 3</td>
<td>.34</td>
<td>1.13</td>
<td>1.49</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Table 4.8 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarter for Cathy’s students.
Cathy’s class spent less time overall studying listening comprehension in the Winter quarter than in the Autumn quarter. The amount of time that she spent on listening comprehension during class sessions, and that she assigned for homework was less during the Winter quarter than the Autumn quarter. Therefore it is not surprising that Cathy’s students did not report spending significant time studying this area in the Winter. (1.2 hours total time Winter vs. 1.73 hours total time Autumn).

The differences in the other tests areas are not as great, and appear almost undetectable in the area of reading comprehension (3.79 total time for the Autumn quarter and 3.77 total time for Winter quarter for reading comprehension). There are some possible reasons that may explain the lack of difference.

For reading comprehension, Cathy increased instructional time during Winter class sessions to focus more on reading comprehension skills. As a result, her students may not have felt the need to devote a significant amount of time to this area of the test. Several students wrote on their questionnaires that they considered class time spent on this topic as preparing them for the test. In other words, students perceived class instruction as reflecting testing demands. They believed that because the teacher devoted instructional time to this area, it was important to the test. One might have expected that students would have increased their study time in this area, but they did not. As they indicated on the questionnaire, they believed that class time counted as study time.

As for the textbook readings, and the grammar and structure portion of the tests, Cathy’s students increased the total amount of time spent studying for these sections in the Winter quarter. Teacher behaviors indicative of washback comprised class instruction that focused more on the textbook readings, and skill based activities that
emphasized the development of reading comprehension and writing. The changes in Cathy’s instructional behaviors appeared to have influenced student behaviors showing increases in the student study times reported, e.g. indicating student washback.

In Cathy’s class, grammar was a focus. For both quarters Cathy admitted that she “loved grammar”, and that she believed it was essential to learning a language. Because her class during the Winter quarter was a night class that met two times a week for two and a half hours each class, the amount of time during one session that could be devoted to grammar actually increased. If class instruction devoted 50-60 minutes to grammar and structure, it is probable that students emphasized this area in their study preparation for the tests. Although the tests did not focus predominantly on grammar, the teacher behavior influenced by Cathy’s preferences, and her past teaching experience, seemed to project the importance of grammar onto the students as well. What is unusual is the students reported that they equated class time spent on reading with test preparation, but their self reported behaviors did not corroborate with the class time spent on grammar instruction.

4.12.5 Study Habits for Diane’s Students

<table>
<thead>
<tr>
<th></th>
<th>AU 04</th>
<th>WI 05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>.58</td>
<td>.12</td>
</tr>
<tr>
<td>Test 2</td>
<td>.18</td>
<td>.23</td>
</tr>
<tr>
<td>Test 3</td>
<td>.23</td>
<td>.22</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>1.30</td>
<td>.77</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>1.28</td>
<td>.90</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>1.25</td>
<td>.81</td>
</tr>
<tr>
<td>Readings from Text</td>
<td>1.47</td>
<td>1.01</td>
</tr>
<tr>
<td>Readings from Text</td>
<td>1.28</td>
<td>.91</td>
</tr>
<tr>
<td>Readings from Text</td>
<td>1.25</td>
<td>.99</td>
</tr>
<tr>
<td>Readings from Text</td>
<td>1.07</td>
<td>.83</td>
</tr>
<tr>
<td>Structure/Grammar</td>
<td>1.07</td>
<td>.68</td>
</tr>
<tr>
<td>Structure/Grammar</td>
<td>1.01</td>
<td>.91</td>
</tr>
<tr>
<td>Structure/Grammar</td>
<td>1.07</td>
<td>.83</td>
</tr>
</tbody>
</table>

Table 4.9 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarter for Diane’s students.
The overall study time for students in all areas decreased from Autumn to Winter quarter for Diane’s students. These changes seem to correspond to the changes in teacher behaviors between the quarters. The decrease in study time spent on listening comprehension appears to be explained by the decreased focus on listening comprehension between quarters. In the Autumn quarter Diane frequently devoted class sessions entirely to listening comprehension, whereas in the Winter quarter none were observed. Listening comprehension activities were still assigned in the Winter quarter, but with less frequency, and in less quantity than during the Autumn quarter.

Reading comprehension and textbook readings were also emphasized less during the Winter quarter than the Autumn quarter, which may have accounted for the reduced amount of Winter quarter student study habits. During the Autumn quarter Diane was frequently observed spending time discussing the textbook readings and using the textbook activities for the readings to expand upon them. Additionally, she would incorporate the grammar and structure instruction with the textbook readings, so that days after the lesson discussing the readings, students would still be “discussing” them within the context of the grammar activities. This behavior, which indicated washback, was observed less frequently during the Winter quarter. This may account for the decrease in the study time. Because Diane did not emphasize the text readings as much during class time, perhaps the students failed to see them as worthy of significant study time. As a consequence, the potential student washback behavior focusing on the textbook readings and contextualized grammar diminished.
Students reported a decrease in the amount of time spent studying for the grammar and structure portion of the test from one quarter to the next. Diane admitted that she focused more on grammar and structure in the Winter quarter, because she felt the students’ skills were weak in this area. Despite the teacher’s increase in instructional time, students reported a decrease in study time. This might be explained by the fact that the students spent so much time working on this area in class, (classes were often 75-95% de-contextualized grammar and structure activities from various workbooks and textbooks, one right after the other), that students did not feel it was necessary to study this area much outside of class. Because class sessions provided so much practice, perhaps students felt that a review of the material was all that was necessary prior to an exam. Or, as Cathy’s students reported, Diane’s students may also have equated class time with study time.

4.12.6 Study Habits for Elena’s Students

<table>
<thead>
<tr>
<th></th>
<th>AU 04</th>
<th>WI 05</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listening Comprehension</td>
<td>Reading Comprehension</td>
</tr>
<tr>
<td>Test 1</td>
<td>.45</td>
<td>.53</td>
</tr>
<tr>
<td>Test 2</td>
<td>.56</td>
<td>.74</td>
</tr>
<tr>
<td>Test 3</td>
<td>.18</td>
<td>.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI 05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>.66</td>
<td>2.70</td>
</tr>
<tr>
<td>Test 2</td>
<td>.38</td>
<td>1.48</td>
</tr>
<tr>
<td>Test 3</td>
<td>.36</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Table 4.10 Self reported student study habits for each test. Average time spent per student studying for each section expressed as a portion of an hour; Autumn and Winter quarter for Elena’s students.
The students in Elena’s class during the Winter quarter increased their overall study time for each portion of the test as compared to the students from the Autumn quarter. The increase was most dramatic in the area of reading comprehension, readings from the text, and structure and grammar. Considering the washback behaviors that Elena exhibited during her teaching, this is not surprising. For example, it was common for Elena to practice listening comprehension in the format of the listening comprehension on the tests. Elena would spend 10-15 minutes focusing on this area, often at the beginning or end of class. Instead of playing a CD of a recording that the students had in their workbooks, she would read a short excerpt from an unfamiliar story in the students’ text. (Each chapter had at least one story that was not part of the required readings for the tests, and she would choose that story). She would write 3-5 questions on the chalkboard in English, and read the excerpt to the students in Spanish, at least twice. The students would have a couple of minutes to answer, and then the class would give the answers. This was a regular occurrence in Elena’s Autumn class, and was even more apparent in her Winter quarter class. This focused washback behavior by Elena, modeling exactly the format of the actual test (reading the selection and having the students answer questions in English), helped to prepare the students for the exam, and demonstrated the importance of listening comprehension skills for the students.

Elena reported in an interview that the biggest change she made in her Spanish 104 class from the Autumn to the Winter was beginning to incorporate specific activities proactively as opposed to reactively. She knew from her Autumn quarter experience what to anticipate in terms of students’ needs, and therefore she didn’t have to wait until the students demonstrated a deficiency. Elena not only continued to exhibit washback
behaviors, she began to exhibit them earlier during the Winter quarter than she did in the Autumn quarter.

The increase in study time in the area of structure might be explained by the fact that Elena spent focused time on developing students’ writing skills, and often grammar and structure became an important part of these lessons. Because she would emphasize different parts of speech and effective ways of writing, such as making transitions, linking ideas together, and using appropriate tenses to express certain ideas, Elena’s students may have compensated by spending more time studying grammar and structure during the Winter quarter. Elena’s instructional behaviors emphasized writing skills from the beginning of the course; she did not wait until the students were required to write in Spanish and then encountered problems. It appears that Elena’s washback behaviors influenced her students by affecting their study habits.

4.12.7 Summary of Student Reported Study Time

Students in all of the classes from the Autumn to the Winter quarter reported changes in their study habits, however the reasons for these changes have not been determined to be only the result of teachers’ increased knowledge of the testing program, or the changes in the distribution of instructional time and class activities. It should also be noted that the groups of students from one quarter to the next were different.

Both teachers and students engaged in washback behaviors during the study. Students seemed to respond to two things: their teacher’s instructional behaviors (meaning what the teacher did in class), and their own increased knowledge of the test demands. If teachers stressed a skill or a concept during instruction, then students appeared to stress that same skill or concept while studying for tests. Students also
appeared to compensate on their own for activities that they believed were important to study for the tests. However, it is critical to remember that there are many factors affecting student study habits, and as some students in this study reported, class time is sometimes equated with study time. Factors outside of teacher behaviors reflecting tests were not measured or considered in this study. The teachers’ instructional behaviors did change from quarter to quarter, and although the degree to which they resulted in washback is unknown, it is apparent that these changes resulted in changes to students’ study habits for the tests.

4.12.8 Final Exam Student Responses; Perceptions of Tests and Instruction

Following the Spanish 104 final examination, students were asked to fill out a questionnaire about their perceptions of and reactions to Spanish 104. The questionnaire was divided into two parts with the first part focusing on questions about the class and the testing program. The second part focused on questions about the different parts of the test, and whether students viewed studying for these parts of the tests as improving their language skills. For the purpose of this study, the questions from Part I of the questionnaire that were tied directly to the Spanish 104 class and the tests, and teachers’ instructional behaviors as they relate to washback, were analyzed. These questions were:

1. The class activities in Spanish 104 were different than those of my previous Spanish classes.
   1  2  3  4

2. The class activities helped me to prepare appropriately for the tests.
   1  2  3  4

3. I studied for the tests by engaging in activities similar to class activities.
   1  2  3  4

6. The teacher/instructor clearly communicated the expectations for the tests.
   1  2  3  4
Students were asked to rate their responses on a Likert type scale ranging from 1=\textit{Not at all}, 2=\textit{Not significantly}, 3=\textit{Somewhat/Slightly}, 4=\textit{Tremendously}. The analyses of the responses included a comparison of the mean student responses from Autumn and Winter quarters for each individual teacher. Table 4.12 represents Anne’s students; Table 4.13 represents Beth’s students; Table 4.14 represents Cathy’s students; Table 4.15 represents Diane’s students; and Table 4.16 represents Elena’s students.

4.12.9 Anne’s students

<table>
<thead>
<tr>
<th>Anne’s class</th>
<th>Mean Response AU vs. WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.7 vs. 2.9</td>
</tr>
<tr>
<td>Question 2</td>
<td>2.6 vs. 3.3</td>
</tr>
<tr>
<td>Question 3</td>
<td>2.2 vs. 2.7</td>
</tr>
<tr>
<td>Question 6</td>
<td>2.5 vs. 3.4</td>
</tr>
</tbody>
</table>

Table 4.11 Mean student response for each quarter for Anne’s students

Students in Anne’s Winter quarter class had a higher response rate than the Autumn quarter class on questions two, three, and six. For instance, students in the Winter felt more strongly that the class activities helped them to prepare for the tests. Students also indicated a slightly higher response rate with regard to studying for tests using activities similar to class activities. This is not surprising given that the instructional activities were noticeably different from Autumn to Winter.

For question six, about students’ perceptions of the teacher clearly communicating the expectations of the test to them, the Winter group had a noticeable difference of opinion as compared to the Autumn group. It was mentioned earlier that during an interview Anne commented to the researcher that she was concerned about the fact that some students in her Autumn class did not feel as though she did an adequate job
preparing them for the tests. Anne’s behaviors reflecting test washback during the Winter quarter made an impression on the students, judging by these responses.

4.12.10 Beth’s students

<table>
<thead>
<tr>
<th>Beth’s class</th>
<th>Mean Response AU vs. WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.5 vs. 2.7</td>
</tr>
<tr>
<td>Question 2</td>
<td>2.6 vs. 2.3</td>
</tr>
<tr>
<td>Question 3</td>
<td>2.0 vs. 2.3</td>
</tr>
<tr>
<td>Question 6</td>
<td>2.7 vs. 2.2</td>
</tr>
</tbody>
</table>

Table 4.12 Mean student response for each quarter for Beth’s students

Only slight differences were revealed in the mean responses between the students in the Autumn and Winter classes for Beth. All mean responses fell into the Not significantly category. For questions two and six there was a decrease in the average mean response between the Winter and Autumn, revealing that students indicated less agreement with the statements during the second time Beth taught the course. The responses suggest that students in the Winter quarter did not feel as though class activities helped to prepare them for the tests as much as the students in the Autumn quarter did. Students in the Winter quarter also did not feel as though the teacher clearly communicated the expectations for the test as much as students in the Autumn quarter did.

In addition to the infrequent washback behaviors that Beth exhibited which may have resulted in these student responses, there are other factors that may also have contributed to the students’ perception. Two of these factors include a significantly larger class size in the Winter section as compared to the Autumn section, and the time
difference of the class between the quarters. The class size in the Winter was more than
double what the Autumn class size was, thus students in the Winter may have had a
harder time getting to know Beth, and/or a harder time making themselves known to her.
The change in time of day that the classes were held may have also contributed to
student’s perceptions of the teacher’s behaviors. The class in the Autumn quarter was at
8:30 a.m., while the class in the Winter quarter was at 11:30 a.m. Educational research
has shown that a change in time of day can affect the teacher’s behaviors, and the
students’ perceptions of the teacher’s behavior. In addition the students’ behavior may be
noticeably different at different times of the day.

Questions one and three revealed a slight increase in response from Autumn to
Winter. The responses indicate that students in the Winter quarter agreed slightly more
than their students in the Autumn quarter that activities in the class were different than
the previous Spanish classes, and that students also related class activities to test activities
by engaging in the same types of activities for studying purposes. Although the
researcher did not observe major differences in the instructional activities between the
quarters, students apparently perceived a slight difference.

4.12.11 Cathy’s students

<table>
<thead>
<tr>
<th>Cathy’s class</th>
<th>Mean Response AU vs. WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.8 vs. 2.4</td>
</tr>
<tr>
<td>Question 2</td>
<td>3.7 vs. 3.2</td>
</tr>
<tr>
<td>Question 3</td>
<td>2.8 vs. 2.5</td>
</tr>
<tr>
<td>Question 6</td>
<td>3.8 vs. 3.6</td>
</tr>
</tbody>
</table>

Table 4.13 Mean student response for each quarter for Cathy’s students
Cathy’s students’ indicated a decrease between the two quarters, implying that students in the Winter quarter felt less in agreement with all of the statements than the students in the Autumn quarter. It is important to remember that in her case, Cathy taught a night class during the Winter quarter and that could have had an effect.

The question that revealed the most noticeable difference in mean responses between quarters for Cathy was question two, and this difference was represented as a decrease. Students in the Winter quarter did not perceive activities as appropriately preparing them for the tests as strongly as students in the Autumn quarter did, although both groups of students did respond in the same category. Again, the importance of the effect that teaching a night class may have had on the teacher and the students should not be discounted. Night classes meet twice a week in the evening as opposed to meeting five days a week during the day. In the researcher’s experience, there are many differences between day and night classes, such as student motivation and discipline, the increased class length, and the demands placed on both teacher and student. Although Cathy exhibited increased washback behaviors in her Winter quarter instruction, it appears that the students did not perceive the change in behaviors in the same way that the researcher was able to.

4.12.12 Diane’s students

<table>
<thead>
<tr>
<th>Diane’s class</th>
<th>Mean Response AU vs. WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.7 vs. 2.6</td>
</tr>
<tr>
<td>Question 2</td>
<td>3.4 vs. 3.5</td>
</tr>
<tr>
<td>Question 3</td>
<td>2.4 vs. 2.4</td>
</tr>
<tr>
<td>Question 6</td>
<td>3.6 vs. 3.8</td>
</tr>
</tbody>
</table>

Table 4.14 Mean student response for each quarter for Diane’s students
The mean response between the Autumn and Winter quarters for Diane’s class differs very slightly, and in some cases the mean response was the same. All of the student responses were within the same categories for Diane’s students across quarters, thus indicating that students’ perceptions of Diane’s behaviors were similar for both groups. The student responses are interesting in that the researcher was able to observe a dramatic shift in the instructional behaviors of this teacher between the two quarters. Diane emphasized non test like activities in her instruction more than she did in the Autumn quarter; however, based on the student responses, it did not appear to affect student perceptions of how well the class instruction and the teacher prepared them for the test.

4.12.13 Elena’s students

<table>
<thead>
<tr>
<th>Elena’s class</th>
<th>Mean Response AU vs. WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.7 vs. 3.1</td>
</tr>
<tr>
<td>Question 2</td>
<td>3.6 vs. 3.4</td>
</tr>
<tr>
<td>Question 3</td>
<td>2.9 vs. 2.6</td>
</tr>
<tr>
<td>Question 6</td>
<td>3.6 vs. 3.3</td>
</tr>
</tbody>
</table>

Table 4.15 Mean response for each quarter for Elena’s students

Question one revealed the most noticeable difference for Elena’s students. The Winter quarter students perceived more of a difference between class activities in Spanish 104 and other Spanish classes than did students from the Autumn quarter. This is interesting because the researcher did not observe an obvious difference in behaviors linked to class activities between the two quarters. However, perhaps because Elena exhibited a frequent amount of washback behaviors during the Autumn quarter, the
perceived difference to the researcher was diminished. Overall, it seemed as though students perceived the class activities as somewhat helpful in preparing for the tests, and that the teacher communicated the expectations for the test in a fairly clear manner.

4.13 Summary Students’ Final Exam Questionnaire

The students in Anne’s, Cathy’s, and Elena’s classes from the Autumn and Winter quarters revealed the most noticeable differences in some of their responses to the questions about the influence of tests in Spanish 104. These are the three teachers who also exhibited the most dramatic differences in their teaching during observations. They also were the teachers who most accurately self-reported these changes in their teaching during the interviews. For Anne, three of the four questions analyzed revealed the greatest differences between quarters of all teachers. Coincidently, the changes that were observed in Anne’s instructional behaviors, and the self-reported changes and discoveries about her teaching, were the most dramatic and measurable of the five teachers in the study. Her Winter quarter behaviors exhibited increased washback that influenced her instruction, and also seemed to influence her students’ behaviors.
CHAPTER 5
DISCUSSION

5.1 Overview

5.1.1 Washback

Testing is an integral part of all education and training programs in academic institutions. As described in Chapter 1, washback is the influence of language testing on teaching and learning (Cheng, Watanabe, and Curtis, 2004). For the purpose of this study, instructional behaviors that included using direct test activities in class, and increased use of skill based activities needed for tests, were considered to be demonstrative of washback. A guiding question in determining whether or not behaviors indicated washback was one posed by Watanabe (2004): “would teaching/learning become different if there were no exams (p. 28)?”

Previous washback research accepts that there are two kinds of washback, positive and negative. Teachers who exhibit positive washback behaviors may incorporate activities into classroom instruction that promote the development of language acquisition, and/or the development of language skills (reading, writing, speaking, listening). An example of positive washback might be a conscious focus on task-based activities that mimic real world situations where students are required to use the target language to successfully complete a task. Task based activities can include, but
are not limited to, writing a letter, gathering information from an authentic text, listening to and comprehending authentic language.

Teachers who exhibit negative washback behaviors may incorporate activities into classroom instruction that inhibit, or do not promote the development of, language acquisition and/or the development of language skills (reading, writing, speaking, listening). An example of negative washback might be a focus on discrete point grammatical activities, and/or avoidance of skill building activities, in order to prepare students for a specific portion of the exam. The teacher behaviors that result in positive or negative washback are the result of the influence of a test. Some additional terms that originated as a result of this study, and that will be helpful to keep in mind while reading this chapter are as follows:

- **background washback**: washback that occurs in anticipation of a test
- **focused washback**: washback that occurs when there is direct knowledge of a specific test
- **prior testing washback**: washback from previous testing experience

Background, focused, and prior testing washback appeared to be evident in this study as exhibited by the participating teachers.

### 5.1.2 Spanish 104 Washback Study

This study was carried out in the intermediate Spanish language classroom, specifically Spanish 104 at the Ohio State University, focusing predominantly on teacher behaviors. The purpose of this study was to determine if washback from required tests occurred in these classes, and how exactly it manifested itself. As a secondary aspect, the study also sought to determine if any student behaviors observed in the Spanish 104
classroom suggested washback. The site was chosen as the location of the study because
the Spanish 104 testing program is considerably different than that of the Spanish 103
testing program. In Spanish 104, tests are largely centered on skills, with approximately
seventy percent of the tests attributed to this area.

The Spanish 104 testing program that served as the principal focus for this study
is not comprised of high stakes tests like those that have been the subject of well known
washback studies conducted by Alderson and Hamp-Lyons (1996), Cheng (2002), Qi
(2004), Shohamy, Donitsa-Schmidt, and Ferman (1996), Wall and Alderson (1996), or
Watanabe (2004). High stakes tests are tests that “provide information on the basis of
which significant decisions are made about candidates, e.g. admission to courses of study,
or to work settings” (McNamara, 2000, p. 133). The testing program in this study
consists of three chapter tests, and one final exam. Other assessments are included when
determining student grades for the course. This is considered a low stakes testing
program because the tests are not the only factor used to determine students’ success or
failure in the Spanish 104 course. Also different than previous washback studies, these
tests were not implemented in an attempt to change teaching, rather they exist as a part of
the curriculum.

The chapter tests account for fifty five percent of the students’ final average.
Therefore it is possible to perform poorly on the tests and still procure a passing grade
(64 or above) in the course. Although this testing program is a low stakes environment,
the tests still play an important role in the academic career of teachers and students.
Tests are a way of measuring student performance and future teaching assignments are, in
part, based on successful student performance and student evaluations.
Tests are also a part of education. Nearly all classes have tests, and all students are expected to perform to the best of their abilities on tests whether it is in a Spanish, Math, Science, or History class. Therefore, teachers and students place significant emphasis on tests despite the stakes.

There were five teachers who participated in the study, all with a minimum of three quarters of teaching in the Spanish language program. For the purposes of reporting the results, these teachers were referred to as Anne, Beth, Cathy, Diane, and Elena (pseudonyms). The students included in the study were those from the teachers’ classes.

The teachers were observed sixteen times over a period of two quarters; eight times their first time teaching the Spanish 104 course in the Autumn quarter, and eight times their second time teaching the course in the Winter quarter. Classroom observations were centered on lessons focused on information that students would be tested on (‘test impacting’ observations), and lessons that did not include information that would be included directly on the test (‘non test impacting’ observations).

Teachers were also interviewed three times during the Autumn quarter, and one time during the Winter quarter. Interviews were oriented to find out a teacher’s perceptions of the demands placed on them by the tests in Spanish 104, and how these tests affected their instruction, especially as the teachers became more familiar with the testing program. Additionally, during the first quarter, after Test 1 and Test 2 had been administered, the teachers completed a questionnaire. This focused on their teaching beliefs, and their perceptions of the impact of the testing program in Spanish 104 on their teaching. All of the data gathering methods sought to collect extensive information about
how tests affected teaching based on teachers’ beliefs, teachers’ past teaching experience, teachers’ knowledge of the testing program, teachers’ instructional behaviors, and teachers’ past testing experiences.

The students in the study teachers’ classes completed questionnaires after each test quantifying their study time for the different sections of that test, and their overall study time for that test. After the Spanish 104 final exam the students completed a longer questionnaire about their overall experience in Spanish 104, and how they perceived their teachers’ behaviors within the context of the testing program, and the Spanish 104 course.

Data analysis revealed that tests affect teachers’ instructional behaviors, and that students seem to respond to changes in teacher behaviors exhibiting washback behaviors themselves. The behaviors suggesting washback exhibited during this study were both positive and negative. The central findings from the study as are follows:

- Tests affect teaching in the Spanish 104 intermediate foreign language classes, but affect different teachers in different ways.
- Increased knowledge of the testing program, and experience teaching a course, can manifest themselves as focused washback in the Spanish 104 intermediate foreign language class.
- Past teaching and testing experience can manifest themselves as a combination of background washback, prior testing washback, and as emerging teacher behaviors in the intermediate Spanish foreign language class, depending on the teacher.
• Positive washback, at the intermediate level, results from knowledge of the present testing program, and ability to effectively teach reading and writing skills in the intermediate Spanish language class.

• Grammar focused behaviors can result from the combination of past teaching experience and washback from prior testing experience.

• Spanish 104 teachers believe they need training to teach skills effectively at this level to prepare students for tests, and to encourage positive washback.

• Teacher beliefs and classroom practices can be in conflict with regard to testing and washback; in the Spanish 104 class teachers do not want to admit that they “teach to a test”, but they want the students to do well on tests, therefore, some modify instruction accordingly.

• Spanish 104 teachers who acknowledge students’ test expectations openly, appear to exhibit more frequent washback behaviors in the classroom.

• Spanish 104 students respond to teacher’s instructional emphasis by exhibiting behaviors suggesting washback.

• Spanish 104 students want to know as much as possible about tests and believe that it is the teacher’s job to inform them.

• Spanish 104 students appear to believe that teachers give significant instructional time and emphasis to materials to be tested; therefore, they adjust their studying accordingly.

• In the intermediate Spanish 104 class, negative student washback can occur when teachers stress topics that are not a major focus of the Spanish 104 testing program.
The remainder of this chapter is organized with a discussion of the central findings presented within the context of the research questions. After discussion of the research questions, the implications and suggestions from the study are presented, followed by recommendations for future research.

5.2 Discussion of the Research Questions

Research questions

1. In a low stakes testing context, what form(s) does washback take? How is washback manifested?

Different forms of washback behaviors were observed throughout the Spanish 104 classes in the Autumn and Winter quarters, and the characteristics of washback changed in these classes from quarter to quarter. One important dynamic of washback revealed in the present Spanish 104 study reflected test anticipation in general. Everyone at this level of schooling has taken a test at some point in her/his life. Due to this life experience, everyone demonstrates some type of washback in all areas that require testing. That is, everyone who has ever taken a test anticipates to some extent what a test will be like: the format, the length of time the test will take, how to study, the material contained on the test, the existing information about the test that one learns from the person giving or writing, the test, etc. This may be referred to as background washback, as it stems from the life experience of taking tests. It is not necessarily specific to the information contained on a particular test; however, because of the general testing experience, people will behave a certain way when they know that a test is imminent. It is logical that these behaviors will all vary according to the person, the purpose of the test, the importance of
the test, and so forth. In this study background washback behaviors included test anxiety, class time discussing details of a test, and increased questions from students about a test.

Once a person is in a particular setting, such as the Spanish 104 class, the washback from a test can become more specific to that test. The test taker learns about the specific material included on the test, perhaps strategies for preparing for and taking the test, and the value of the test within the context of the course and the test taker’s academic career. This may be referred to as focused washback, because it specifically focuses on the precise test. For the Spanish 104 study, focused washback included increasing activities centered on listening, reading and writing skills, the textbook readings (direct test activities), and to a lesser extent, contextualized grammar and structure topics from the course textbook. The class activities that concentrated on these topics helped to prepare students to successfully meet the demands of the test. Therefore, they were considered washback because they promoted the skills that were directly tested.

Washback was evident in a number of teacher behaviors observed in the classroom during this study. Table 5.1 depicts the washback behaviors that were identified in the research, and the teachers in the study who exhibited those behaviors most during the Autumn and Winter quarters.
<table>
<thead>
<tr>
<th>Teacher Behavior</th>
<th>Washback</th>
<th>Teachers Autumn quarter</th>
<th>Teachers Winter quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading and writing from text activities</td>
<td>Spanish 104 Test activity</td>
<td>Beth, Diane, Elena</td>
<td>Anne, Cathy, Elena</td>
</tr>
<tr>
<td>General skill building in reading and writing</td>
<td>Spanish 104 Test activity</td>
<td>Diane, Elena</td>
<td>Anne, Cathy, Elena</td>
</tr>
</tbody>
</table>

5.1. Identified washback behaviors, and the teachers who exhibited them most during the Autumn and Winter quarter.

These behaviors were identified as a result of observations, teacher interviews, and teacher questionnaire analyses. During the Autumn quarter teachers were observed incorporating test activities in class that included activities based on listening comprehension, the required readings from the text, and grammar and structure. Some washback was evident, but the degree to which it was observed, varied according to the teacher. A large portion of instructional time was centered on grammar, especially in Anne, Beth, and Cathy’s classes.

Because all of the teachers in the study had previously taught in the elementary level program prior to teaching Spanish 104, their washback behaviors largely reflected their previous teaching and testing experiences, especially that of Spanish 103. During the Autumn quarter, it was quite common to observe teacher behaviors that reflected experience with a course and a testing program focused predominantly on grammatical accuracy with a lesser emphasis placed on skill demonstration. These characteristics are the major components of the course and the testing program for the Spanish 103 course. However, the more familiar some of the teachers became with the tests in Spanish 104, the more they adjusted their instructional emphasis to accommodate the demands of the Spanish 104 tests (based mostly on skills), and the students’ need to meet those demands.
During the Winter quarter, as the teachers’ knowledge of the Spanish 104 testing program increased, and as teachers became more aware of the students’ need to have information about the test to be able to prepare effectively for the tests, teachers’ washback behaviors changed. Most teachers still included activities they used during the Autumn quarter, but more attention was given to the required textbook readings, and the skills necessary to read and write effectively in the target language. Teachers’ behaviors changed to different degrees. Those who demonstrated the most washback in Winter quarter were Anne, Beth, and Cathy, as they used the textbook reading activities extensively, and Elena who even went beyond those activities to spend focused time developing students’ skills in reading and writing. The teachers were influenced by the tests in Spanish 104 and they altered their teaching accordingly. They devoted more class time to skills needed for the tests, and made direct test like activities a regular part of class.

During the Winter quarter, Beth and Diane did not place more emphasis on skill development; rather, they shifted their behaviors to emphasize grammar instruction. Their washback behaviors from the Spanish 104 tests were limited, and they appeared to demonstrate some prior testing washback behaviors from previous testing experiences. Their predominant class behaviors focused on grammar and structure. The reason that these teachers gave for focusing on grammar was that they felt grammar was a particularly challenging area of the test for students, and the students needed to improve their abilities in this area. Beth and Diane’s behaviors did not consistently focus on skill development, or the most heavily weighted parts of the current Spanish 104 tests. Their behaviors appeared more focused on previous Spanish 103 course components and tests.
Consequently, their washback behaviors during the Winter quarter can be considered negative because they concentrated on test components that were not a main focus, and time was shifted away to areas that were not at the heart of the course objectives. This instructional shift took time away from skills and current test material.

Anne, Cathy, and Elena were observed giving special attention to writing skills in the Winter quarter, demonstrating a noticeable increase from the attention given to writing skills in the Autumn quarter. This was considered focused washback behavior because it directly reflected the demands of the Spanish 104 tests, and provided opportunities for students to practice these skills to prepare for the tests. Teachers understood the content and the format of the test, and consequently they modified their instructional behaviors to include more skill building activities, and more direct test like activities into their Winter quarter classes.

Anne, Cathy, Diane, and Elena demonstrated positive Spanish 104 washback, and made true progress in developing students’ skills via skill building activities. These teachers also had the most experience teaching, and/or some educational training or background in teaching or pedagogy. For example, Anne made the most significant and dramatic changes in her teaching, and also self reported these changes in interviews. Prior to Autumn quarter, she had just completed a Master’s degree in Foreign and Second Language Education. In her educational program she learned about techniques and methods for teaching reading and writing, and skill development. Based on her Autumn instructional behaviors she did not recognize that these were the skills she needed in the Spanish 104 classroom. However, before the Winter quarter began she made the connection and then dramatically adjusted her teaching. Her Winter quarter behaviors
reflected significant skill building. These behaviors were indicative of washback because she spent significant time on direct test like activities, and teaching content directly related to the test.

Cathy was an experienced teacher with nine years of teaching experience, and a doctoral student in Spanish literature. She had knowledge of the current educational literature, and pedagogy in foreign language education. She attended and presented at foreign language education conferences, and maintained a keen interest in teaching. Cathy frequently referred to the influence of Madeline Hunter, an educational researcher, as she reflected upon her teaching. In classroom sessions she was observed talking about this educational researcher to her students. During the Winter quarter, Cathy exhibited more focused washback behaviors that included increased instructional time on required textbook readings, more focus on skills, and the use of more direct test-like activities during class sessions.

A deterrent to the positive washback behaviors influencing Cathy’s teaching was her passionate belief that grammar was the cornerstone of language learning. This fact may be attributed to both her past teaching experience, and her literature training. She viewed grammar as essential to correct and effective interpretation in literature. Changing the tense or meaning of a word can affect the interpretation of something. Additionally, if Cathy’s past teaching experience occurred in foreign language programs that emphasized grammar (as many do), then it is likely that this was why she often made grammar an important focus of lessons. Although her experience for the most part was an asset to her teaching, it appeared that it in some ways it may have resulted in prior testing washback, and negative washback behaviors focused on grammar. The
instructional behaviors that Cathy exhibited focusing directly on grammar instruction, and in her case appeared to be more related to her personal beliefs than her experience in the Spanish 103 course. Although it is impossible to distinguish how much of her behavior was related to past teaching and testing experience, and how much was related to her personal preference, in her situation, it looks like her preferences were the stronger influence. Considering the comments she made during her interviews, and the observations of her class, there is no doubt that this teacher believed that grammar was essential. Therefore, her belief had a strong influence on her practice.

Elena, a teacher with twelve years of experience, did not make dramatic instructional changes between quarters. Rather, she was the most consistent from the Autumn quarter in focusing on students’ skills, and on developing them. She used her own experiences as a struggling language learner to inform her instruction and help her students. She had past experience teaching reading and writing, and she enjoyed teaching both reading and writing. Because she struggled at one time with reading and writing in a foreign language herself (she is a native speaker of Spanish), she was often able to anticipate her students’ language problems. Interviews revealed that her own language learning also helped her to offer a unique perspective that none of the other teachers could offer because her reasons for learning another language were different than the teachers, and the Spanish 104 students. She learned a foreign language to function successfully in another country, not to meet a language requirement, nor for folly (travel, cultural experience, etc.). As a result of her experiences, her class instruction focused more on skills and less on the components traditionally stressed in the typical American foreign language programs that were familiar to both Spanish 104 students and other
Spanish 104 teachers. Her experiences had also made her aware of the importance that students gave to tests, and her instruction reflected this understanding as well. Washback in her classroom was positive, and from the beginning was more focused reflecting reading and writing skills for language learning, and success with Spanish 104 tests. Elena admitted that she wanted the students to learn the language not only to be successful in class, but her opinion was that learning another language was valuable.

2. What characteristics and factors draw distinctions between washback and teacher behaviors that reflect adjustments to the course as opposed to test-related washback? How can these two domains be distinguished?

Washback behaviors observed in this study were similar to those described in high stakes testing washback research. Cheng (2002) reported teachers changing their instructional behaviors to include test like activities that come from textbooks, and to focus on skills. Watanabe (2002) described washback behaviors that stemmed from “teachers’ concerns for students’ proficiency levels”, where teachers used grammar activities from past exams because they felt that students were weak in certain skill areas (p. 139). Ferman (2002) explained how washback was observed through an increased focus on skills included on the test, resulting “in an increase in time allotted for the development of these skills” (p. 204). Although the context of the present study was low stakes testing, many of the teacher behaviors were the same. Similar to Alderson and Hamp-Lyons (1996), this study also found that tests affect what and how teachers teach, but the “effect is not the same in degree or in kind from teacher to teacher…(p. 295)”.

More noticeably during the Winter quarter, once the teachers were more experienced with the course, and more confident with the curriculum and materials, it
was common for their behaviors to include direct test-like activities such as writing summaries of the required readings from the text, answering questions about these readings sometimes in groups, and sometimes comparing answers to the Preguntas de Orientación. In addition, some teachers incorporated activities that focused on listening comprehension. At times listening comprehension took the form of class activities focused on a video from the text, and other times they utilized the listening comprehension tracks that both teacher and student had access to in their workbook. In other instances, activities modeled the test directly. Elena practiced listening comprehension by reading an excerpt from a selection in the text, and then the students answered questions that were written in English on the board.

Class observations during the Winter quarter revealed many more direct test-like activities overall than class observations from the Autumn quarter. Teachers knew more about the test expectations, and about the students’ need to understand and perform well on the tests. They demonstrated more concern with regard to student performance on the tests and student’s attitudes toward the class. As a result of this, instructional time reflected test activities and test expectations more than in the previous quarter.

As is expected, the increased instructional time during the Winter quarter on skill development also appeared to be influenced by an increased familiarity with the course. Obviously as teachers learned more about the Spanish 104 course and became more familiar with the curriculum, they had a greater understanding of the material. This understanding, coupled with the fact that they had taught the material to a group of students once before, helped teachers to anticipate the potential problems for students the second time around. This allowed them to plan instruction accordingly, and to emphasize
areas that were problematic for students in the past. However, increased familiarity with
the course content was not the only factor that appeared to influence instructional
behaviors. Teachers spent increased instructional time on test content material. The
tests demanded that students demonstrated skills, with the most significant difference
between Spanish 103 tests (familiar to the students) and Spanish 104 tests (new to
students) being the amount of writing in Spanish, and the questions about the textbook
readings. Not only did teachers spend increased time in these areas, teachers specifically
included direct test like activities as a regular part of the class. They also spent
considerable time discussing the test and working on activities with the purpose of
preparing students for the tests. Although the teachers were more familiar with the
course expectations and the material, they also had a keener sense of the test
expectations, and the importance of the test to the students.

Anne, Cathy, and Elena, recognized the link between skill development and test
preparation, and their instructional behaviors reflected this. It was common to find each
of these teachers promoting reading by teaching their students reading strategies, and
having them engage in activities that encouraged them to use their reading skills. These
activities included reading new texts, talking about a text as a class, activating prior
knowledge about a text prior to reading and interpreting it, or even listening as the
teacher read it aloud, acting it out for the students.

These three teachers also promoted writing in their classes. Activities that
promoted writing included sharing writing samples and analyzing them as a class, writing
together in groups, and modeling effective writing for the students. By instructing the
students how to write more effectively in Spanish, Anne, Cathy, and Elena not only
helped the students to improve their communicative abilities, they also helped them to prepare for paragraph length responses for test essay questions.

Wall and Alderson (1996), and Alderson and Hamp-Lyons (1996) described negative washback behaviors that occurred when teachers eliminated, or reduced instructional time, for activities that were not directly tested. In this study negative washback behaviors observed included the elimination of, or diminished frequency of paired oral activities in three of the five classes (Anne, Cathy, and Elena), and the diminished frequency of teacher generated post learning activities (Diane). Because teachers did not believe that these activities were useful in preparing students for the exams, they ceased to incorporate them into their class activities. The fact that the activities might have still been useful for skill building was not considered.

For example, in the Autumn quarter, Cathy, included summarizing activities modeling the chapter theme at the conclusion of a chapter. She incorporated a lip-synch activity at the end of the Chapter 1, but she later abandoned this activity Winter quarter. The reasons Cathy gave for not using this activity during the subsequent quarter were mostly related to limited time within the fixed schedule. However, Cathy did include more direct test like activities during that quarter, thereby demonstrating more focused washback behaviors on test material. Although she attributed not using the activity to time constraints, she found the time to include more test-like activities. Her instructional behaviors were more associated with test washback than with increased familiarity with the course.

Washback behaviors different than those reported by Watanabe (2002), but seen frequently in this study, were those that came from teachers’ past teaching and testing
experiences. These behaviors were directly related to teacher beliefs, something that Watanabe (2002) described, but not within this context, as his study focused on high stakes testing. Nor does Watanabe describe these behaviors in relation to teachers’ past testing experiences. For this study, the Spanish 104 test was a component of the intermediate language course that was new to the teacher and the student. Although it represented a dramatic testing change in both form and content, the teachers were most familiar with the previous testing program from the elementary level program where the emphasis on structure and grammar is fairly intense. Teachers’ behaviors exhibited during Autumn quarter stressed the area of grammar even during instruction in Spanish 104 indicating prior test washback from Spanish 103. During Autumn quarter it was common to see prior test washback behaviors in Anne, Beth, and Cathy’s classes as considerable instructional time focused on discrete point grammar activities. In an interview, Cathy said that when her students requested more focused time on areas directly related to the tests (the required readings from the text and reading skills) she found it hard to believe that she was not spending sufficient time on this material. Her perception was that she was teaching it effectively, but she viewed other areas as more important. Her past testing experience was an influential factor on how her behaviors indicated washback in the Spanish 104 classroom.

The behaviors resulting from past teaching and testing experience during the Winter quarter manifested themselves as prior test washback in Beth and Diane’s classes. Both of these teachers anticipated the grammar sections from Spanish 103 tests. Watanabe (2002) described teachers using activities as a result of concerns for student’s limited skills, which was exactly what these two teachers did in their classes. Both Beth
and Diane saw students’ grammar skill as particularly weak, and therefore they focused instruction intensely in these areas, although the Spanish 104 tests didn’t have a predominant focus on this area.

Another teacher behavior not described in the high stakes testing washback literature came from students’ expectations regarding a test. As mentioned, the teachers in the Spanish 104 study exhibited focused backwash behaviors that manifested themselves in the use of direct test-like activities, focused practice for the test, and/or information about test content and format. One of the contributing factors to these teacher behaviors was students’ testing expectations. All teachers acknowledged openly during Autumn quarter interviews that their students had clear expectations about tests. The students had communicated to their teacher either through class surveys, class discussions, or one on one sessions, that they expected her to inform and prepare them for the test. The researcher confirmed these student expectations with data from the student questionnaires. Because these teachers listened to their students, and they accepted that their students’ needs were important, they made efforts to provide the students with increased information about the tests. This quantity and type of the information ranged from providing format information about the set up of the test, and the types of possible activities that students could expect to see, to the content of the test material, practice test-like activities, and strategies for effective studying. During the Autumn quarter, all teachers provided information about the format of the test, and even the possible types of activities, but during the Winter quarter Anne, Cathy, and Elena demonstrated increased focused washback behaviors for the tests.
One of the questions that this study has raised is: how do the students’ testing expectations interact with the teacher’s past test experiences? It appears that both influence instructional behaviors in the classroom, but to what extent are teachers influenced more or less by their previous testing experiences, as compared to the influence of their students’ present testing needs? Although this study cannot answer this question, future research investigating this relationship is warranted.

3. What similarities and differences exist with respect to teacher instructional behaviors and washback based on teacher perceptions, student perceptions, and the researcher’s perception?

Instructional strategies were visibly altered as teachers became more familiar with the testing program again to different degrees. Anne and Cathy actually incorporated more test-like activities into their teaching, and began “training” students to successfully perform the activities needed for the tests. Diane and Elena altered their instruction in more subtle ways by including more skill building activities focusing on developing students’ skills in the areas of reading and writing, skills emphasized on the exams. Beth, while not making dramatic changes in her instruction toward more test-like activities, appeared to make changes in the information that she communicated to students prior to exams. She informed them of the format, and the types of questions and the material that they might expect to see on the tests. For Anne, Cathy, Diane, and Elena, once they knew about the tests, they each modified their practices in some ways to better prepare the students for the test. The tests affected the teaching in the intermediate Spanish language classroom resulting in an increase in teachers’ washback behaviors during the Winter quarter.
As the instructional time totals from the observations indicated, some of the teachers made more observable changes in their instructional behaviors between quarters. These changes seemed to be related to two things: the teacher’s perception of the course and the tests, and the students’ perception. After the Autumn quarter Anne, and Cathy better understood what the tests demanded from the students, and they made the most notable instructional differences, realizing that their previous instruction wasn’t necessarily matching these demands. For the Winter quarter these two teachers acknowledged that students needed more instruction in the area of reading, and they created opportunities to help students read more effectively in the target language, and gain a better understanding of the testing material.

During the Winter quarter, Anne, Cathy, Diane, and Elena all incorporated more pre-, during, and post-reading activities into the class sessions that focused on the required readings. They often spent more time on the readings than what the calendar allotted, and they sought ways to involve the students in the reading (acting out the stories, writing different types of comprehension questions about the stories, writing about the stories together, etc.)

Beth did not perceive a gap between her instruction and the demands of the tests between quarters the way her colleagues did, and therefore she continued with mostly the same instructional techniques that she previously used. However her frustration was apparent because she commented, “it is frustrating that the Preguntas de Orientación are so different from the tests, because the students think that that’s what they’ll need to answer on the test, and then it’s completely different”. It never became clear to her that the textbook questions were a starting point, and she had to manipulate instruction to
focus on developing reading abilities. Her behaviors continued to focus largely on what she was used to doing in her previous teaching experiences which was grammar and structure. This reflected prior testing washback. The link between the curriculum and the testing program never became obvious to her.

The second factor influencing changes in the instructional behaviors of the teachers was related to students’ perceptions of the class and the tests. For teachers, student performance exposed their needs just as much as students telling teachers what their needs were. Anne and Cathy’s classes specifically told their teachers that they needed more support with reading and writing. The students commented on the difficulties that they had with the transition between the Spanish 103 (elementary) and Spanish 104 (intermediate) courses. Because the students brought this to the attention of their teachers, even though it was the middle or end of the first quarter, Anne and Cathy reported that they began to make changes during the remaining weeks of the Autumn quarter. Anne and Cathy knew before they began the Winter quarter that students needed support with reading and writing. Anne specifically commented, “I assumed that they [the students] knew how to do this, I was wrong”. The teachers observed weaknesses as reflected in the students’ performance, and the students acknowledged their weaknesses and their needs as well directly to their teachers. During the Winter quarter, both teachers made observable changes based partially on student concerns.

Although Elena did not make dramatic changes to her teaching during the Winter quarter, because she already included significant skill building and test like activities during the Autumn quarter, she began to implement the skill building and test activities much sooner in the Winter quarter. She did not wait until the students asked for help, or
the students demonstrated difficulties, rather she was able to anticipate. Her washback behaviors incorporated skill building and test preparation activities from the beginning of the course.

The second component of the observed teacher behaviors related to students’ perceptions, resulted from the students’ performance on the test. Anne, Cathy, Diane, and Elena were observant of their students. They realized that often student performance was affected by more than simply the amount of time spent studying for a test. They noticed that there were certain tasks that the tests required of the students that they were unable to do successfully such as answering questions about the textbook readings, or writing comprehensible answers completely in Spanish. These four teachers then spent specific time instructing the students in reading and writing, so that when they came to the test, they would be able to meet the demands of the test more effectively. Anne, Cathy, Diane, and Elena altered the instruction in the classroom to meet the students’ needs and to help prepare them for the test. The teachers exhibited washback as a result of knowledge of the testing program, and acknowledgement of students’ perceptions about the test.

Although Diane did spend time on reading and writing in order to help her students prepare for the tests, her instructional behaviors changed quite unexpectedly during the Winter quarter. Beth and Diane’s instruction changed to include greater emphasis on grammar and structure. This change appeared to come from their interpretation that poor student performance on tests was a consequence of weak skills in grammar. Therefore, these two teachers believed that the students would read and write better if they could master the grammar. Diane commented, “I can’t believe that they’re
making these mistakes at this level- it’s terrible. That’s why I really focused on the grammar this time around.”

Beth also held students accountable for grammatical accuracy on the test, taking off points for grammar errors. Because this caused students to view grammar as one of the most important aspects of the tests, students appeared to spend less time focused on other areas of the test. Beth’s behaviors encouraged negative student washback behaviors as well, because students were afraid of the consequences of poor grammar on the tests. Both of these teachers’ perception of the test, and their students’ performance on the test directly influenced the changes that Beth and Diane made to their instructional practices once they learned more about the exams. The washback behaviors that Beth and Diane exhibited reflected a limited emphasis on skills during instructional time in class sessions for the Winter quarter.

Reasons that may account for teachers not making notable modifications to their instruction to reflect test expectations, most likely include limited expertise. Teachers’ limited perceptions and their interpretation of students’ performance forced them to focus on the areas that they thought were important, or that students should know. Watanabe described how teachers overemphasized certain skills based on their own opinions (2004, p. 140). He further went on to say that “the present research suggests [the]….need to incorporate a course in which teachers are provided with a variety of teaching methods involving various test tasks (p. 141). Beth and Diane already had one significant advantage in teaching reading and writing skills to their students; their own language learning. Although all teachers’ responses on the questionnaire revealed that they
considered this to be very important, manifesting this in classroom instructional behaviors was an altogether different issue.

As stated, Beth made the least amount of changes in her instruction between quarters. A Spanish literature student with limited experience teaching, but significant experience in reading, writing, and interpreting Spanish, her own language learning could have provided a resource for her teaching as she was essentially teaching her students the same things, albeit at a different level. However, she never incorporated this into her instruction. Important to note here as well is that she was the one teacher in the study without any background in Education, or training in teaching or pedagogy. Diane, who made instructional changes that reflected a shift away from the reading, writing, and skill activities she used in the first quarter, and emphasized grammar and structure for the second quarter was an experienced teacher. However, she became focused on the students’ deficiencies thus devoting her efforts to these areas. She interpreted the weak grammar abilities as a call to focus more intensely on explicit grammar for future sessions.

Aside from the amount, form, and type of washback seen in this study, it is important to note that all teachers in the Spanish 104 study wanted their students to do well on the tests. When students do not do well on exams, they often blame the teacher. Then students give poor evaluations to the teacher. If students perform well on a test, especially a test that is skill based, there is a reasonable assumption that students have learned the material; thereby making it easier for future instruction and learning. A teacher who spends significant time preparing the students for the exams, and giving them practice type activities is not as likely to receive the blame for poor student
performance. Data analysis in this study indicated that at times, teacher beliefs and classroom instructional practices were in conflict regarding testing and washback. Anne and Cathy specifically made comments such as “I don’t teach to the test”, or “not to say that I teach to the test because I don’t”, however, both of these teachers wanted their students to do well on the tests. Consequently, they modified their instruction to facilitate students’ test success accordingly during the Winter quarter.

Students also want to do well on tests in spite of the tests’ impact, and regardless of their interest in the subject matter. In the case of Spanish 104, all students need to pass the course to fulfill the language requirement for graduation. Therefore it was in the teachers’ and the students’ best interest to work toward success on the tests, regardless of the effect on language skills and abilities. Hence there was a noticeable increase in washback behaviors during the Winter quarter.

4. What elements of the intermediate testing program encourage positive washback?

The intermediate Spanish 104 language testing program encourages mostly positive washback. The elements of the testing program that encourage positive washback are the skill-based activities that are part of the tests. The majority of every test focuses on listening, reading, and writing skills (approximately 70 percent of the exam). To be successful on the test, the students need to have well developed skills in listening, reading, and writing. They must be able to read, interpret, and summarize in the target language, and they must be able to communicate their ideas effectively through writing in the target language. Therefore this encourages skill development in these areas, i.e., positive washback. A concern, however, is that the intensity and degree of positive washback appears to be directly related to the teacher. The extent to which students
actually develop these skills is dependent largely upon the teacher, as she determines the course of classroom instruction. As the teachers in this study demonstrated, if a teacher chooses to focus instruction on other aspects of language learning besides skills, students are left to develop their listening, reading, and writing skills on their own. The teacher who recognizes the importance of these skills not only in language learning, but also as a component for successful test performance, has the ability to exploit these skills in the context of the class instruction. This in turn will help the students to better achieve success on the tests, and to better develop their language skills. This is a primary example of positive washback.

The way that Spanish 104 integrates the testing program with the curriculum also encourages positive washback. The textbook provides ample opportunities for working with the stories, including a variety of reading and writing activities ranging from basic to more complex (interpretation, speculation, prediction). The textbook also contextualizes the grammar and structure activities within the theme of the chapter, offering the teachers an opportunity to exploit the readings through the study of the grammatical concepts.

5. What elements of an intermediate testing program encourage negative washback?

This study indicates that some elements of the intermediate testing program contribute to negative washback. The fact that certain types of activities are not included on the tests encourages negative washback. Teachers cease to use the activities in the classroom because they are not on the test, and therefore teachers do not view them as helpful in preparing for the tests. For example, because there is no oral component on the exam, many of the teachers utilize limited class time for oral practice activities. The emphasis on oral skills typically appears in the form of class discussions about assigned
homework. The class discussions focus on the required readings from the text, and/or the activities that accompany the readings; both topics included on the test. But, students do not frequently engage in extended discourse, nor do they frequently engage in activities designed specifically to build their oral proficiency skills. Because teachers do not spend significant time on oral proficiency skills, the improvement in the students’ oral skills can be limited for students. Oral work in Spanish 104 consists mostly of the paired grammar activities that students complete, partially in English and partially in Spanish, and stresses the formation of sentences, or the correction of grammar homework. As seen in this study, the teachers’ washback behaviors tend to stress other areas that are emphasized on the tests.

Another factor that contributes to negative washback is the volume of material that must be covered during the scope of the course. Teachers in this study had innovative ideas for activities that could further develop students’ presentational skills, and would give students opportunities to interact orally in more sophisticated and complex situations with their classmates. However, the fixed schedule allows little time for the inclusion of extra activities developed by the teacher. Although there is flexibility with the schedule in terms of the order that activities are completed, the tests and other assessments are pre-determined and teachers must adhere to those dates. This was particularly evident in this study during the Winter quarter of teaching. Teachers followed the schedule much more closely during the Winter quarter than they did in the Autumn quarter. Perhaps they sensed that the timing of the lesson is planned to provide students with the instruction and some practice opportunities in order to comprehend the material included on the tests. Although adhering more to the precise planning of the
calendar allowed for more positive washback activities as teachers used more test like activities, it also offered limited time for teacher generated activities that could be used to build skills.

6. Which of the student behaviors previously reported in the literature as indicating washback, were observed in the classroom? What additional behaviors suggesting washback were observed?

Wall and Alderson (1993), Alderson and Hamp-Lyons (1996), Cheng (2002), Watanabe (2004), Shohamy, Donitsa-Schmidt, and Ferman (1996) have all conducted washback studies that have focused almost exclusively on teachers’ behaviors. These well known high stakes test studies used teachers as the primary participants in the study, and therefore their focus on teacher behaviors is logical. Ferman (2004), however, used students as part of his sample in his study about the washback of an EFL national oral matriculation test. In that study, students reported that they “compensated for not having learned for certain parts of the tests in class by learning on their own” (p. 205), and they also reported that the material tested did have an effect on the importance of the material to them. In the Spanish 104 washback study, evidence of student washback from student questionnaires revealed that students adjusted their study time according to the focus of teacher’s instructional behaviors in the classroom. The study habits reported reflected two scenarios: studying more for topics and activities that were stressed during class, while reducing study time for topics and activities covered less, or eliminated from class; or studying less for topics and activities that were stressed during class because some students perceived the added class coverage as fulfilling the obligation for study.
Students also demonstrated washback behaviors during class when they participated in test-like activities, and skill building activities with ease.

Some students who perceived their teacher’s lack of emphasis on skill based or test-like activities as a sign of diminished importance for the activities, chose to place less emphasis on them, by consequently focusing less on these areas in their preparation for tests. For other students, in classes where test-like activities were not a significant part of the teacher’s instructional repertoire, students compensated by increasing study time in these areas on their own. Because there was a limited focus on these particular topics, the students appeared to sense a need to focus more intently on activities to build these skills in order to prepare for the tests. For most students, decreased instructional time by the teacher appeared to result in negative washback as decreased study time on the part of the students.

Students in this study from Cathy and Diane’s classes commented on their student questionnaires that they viewed classroom activities as partial preparation for tests, most likely because instruction often focused significantly on direct test-like activities. These students may not have frequently engaged in activities outside of class that would be suggestive of washback, but their in-class behaviors indicated washback. These students reported affording test-like activities less study time outside of class, most likely because their teacher made them part of the class instruction.

Judging from the self reported student responses about study habits for each test, and the comparison of these responses from the Autumn to the Winter quarter, it is evident that student washback behaviors were mostly related to the teacher washback behaviors. In other words, students in Anne and Elena’s classes, where reading and
writing skills were emphasized more significantly in the Winter quarter, spent more time studying these areas. The increase in the amount of instructional time devoted to specific areas by Anne and Elena indicated an increase in the amount of study time by the students in these same areas. These students perceived that what their teacher emphasized during class was important for them to study in order to be prepared for the test. Although other reasons may have also contributed to increased or decreased study time, it seems apparent that teachers’ behaviors were influenced by tests, and that corresponded by affecting students’ behaviors.

Likewise, material and/or skills that were emphasized less during class, appeared to receive less study time from the students. For example, in all teachers’ classes, there was a reduction in the amount of instructional time focusing on listening comprehension skills during the Winter quarter as compared to the Autumn quarter. Consequently, students also reported a corresponding decrease in their study times devoted to this area when preparing for tests. Students’ questionnaire responses indicated that they perceived that listening comprehension was not as important as readings from the texts, and writing, because they spent less time studying in this area. Therefore, students’ washback behaviors in part manifested as decreased study time spent on this activity.

Students’ responses from test and final exam questionnaires supported the idea that students will engage in washback behaviors on their own, based on what they feel as though they need to know for a test. (Remembering that what they feel they need to know is largely influenced by the teacher’s behavior). In Anne, Cathy, and Elena’s Winter quarter classes, where test-like activities were common, students reported engaging in activities similar to class activities to prepare for the tests. Teachers included
mock test-like activities in their classes to inform the students of the tests, and to help the students prepare for the tests. The students reported that they perceived the class activities as helpful, and therefore mimicked them while preparing for the tests on their own.

In Spanish 104, it is the teacher who communicates the majority of the information about the test to the students. Students can find out test information from other students, from past Spanish 104 tests, or on the Spanish 104 website, but the teacher serves as the primary source of information. Teacher interviews, classroom observations, and student questionnaires, revealed that students want to know as much as possible about the test, and they believe it is the teacher’s job to communicate that information to them.

An interesting question that this study raised is whether teachers realize the power they exert over the potential washback behaviors of the students simply through their own instructional behaviors. If teachers were aware that students perceived that what is presented through direct instruction during class sessions is necessary knowledge for tests, then teachers might be more conscious about their choices, and what they choose to emphasize during instruction. In this study, the data analysis revealed that student washback behaviors reflected teachers’ instructional behaviors. This occurred when the teacher provided information about the format of the exam and the material to be included on the test, and when the teacher made conscious decisions to focus instruction on specific areas, and to include specific activities in class sessions. When teachers focused on reading and writing skills more during the Winter quarter, overall students engaged in more reading and writing behaviors to help to prepare them for the tests. If
teachers were more conscious of the influence of their instructional behaviors on students’ behaviors, they may be more able to influence the washback behaviors in their students.

Bailey (1996) has described student washback behaviors that include,

practicing items similar in format to those on the test; studying vocabulary and grammar rules; reading widely in the target language; applying test-taking strategies; and requesting guidance in their studying and feedback on their performance (p. 264-5).

In the Spanish 104 washback study, student washback behaviors observed in the classroom mostly included engaging in test related activities such as writing summaries, writing various types of comprehension questions, writing about the characters from the required reading, and increased use of the target language.

The observations during the Winter indicated that, in Anne, Cathy, and Elena’s classes where these test-like activities became a regular part of typical class activities, the students would move into the activities quickly, and began participating immediately. The students were comfortable engaging in activities focused in these areas. They also demonstrated more frequent use of the target language for classroom activities, as well as some classroom exchanges prior to the start of class. On occasion students spoke to the researcher in the target language. Skill development was a focus of these classes, and consequently students’ behaviors reflected this.

In Beth and Diane’s classes during the Winter quarter, where these types of activities did not constitute a regular part of typical class sessions, it was common for students to hesitate about engaging in the activities. For example, in their classes, students indicated concerns about their understanding of the required readings that they
would voice prior to class sessions, and after class sessions with classmates, and with the researcher. Additionally, if Beth or Diane asked the students to form groups to participate in activities that focused on reading and writing, many students would spend time after the teacher gave instructions, asking each other, and sometimes the researcher, about what they were supposed to do. Or, students would frequently ask the researcher for help or feedback with a particular activity. Typically, students in Beth and Diane’s classes conducted their work in English, with very little Spanish, except when the teacher came to their group to check on them. It was apparent in these classes during the Winter quarter that students were not accustomed to frequent test-like activities involving reading and writing. Classroom observations confirmed that these activities were not a focus of regular class instruction, and consequently the students’ reported behaviors did not reflect skills. Furthermore, students in these classes demonstrated limited use of the target language for communication.

It is important to reiterate that the extent to which student behaviors can be said to reflect only test expectations is limited due to the variety of other factors influencing student behaviors that include but are not limited to, student motivation in general, student motivation for and interest in taking Spanish, rapport with teacher and classmates, classroom instruction, student perception of the importance of the test, student perception of the demands of the test, and student perception of classroom activities as reflective of testing expectations. Still, data from this study indicated that students did change their habits between quarters in manners that corresponded to the changes in teachers’ instructional behaviors. Teachers influenced students’ washback behaviors in positive ways when they increased instructional time on material directly related to the test, and in
negative ways when they increased instructional time on materials that were not a major focus of the test (such as grammar), or when they eliminated activities from the class because they were not material tested directly such as oral skills.

5.3 Implications and Suggestions Arising From the Study

Spanish 104 is an important course for both students and teachers. For students, the Spanish 104 course is a required course for those who elect to study Spanish, and need to fulfill the language requirement for graduation at the Ohio State University. For many of the teachers in the Department of Spanish and Portuguese, success or failure in teaching this first level intermediate language course is a consideration that often determines the courses that a teacher will be assigned to teach in the future. Because the testing program is a major difference between Spanish 104 and Spanish 103, understanding the teacher behaviors that are influenced by and related to the tests, is paramount. In doing so, it is possible to influence washback behaviors of teachers and create programs that prepare them to become successful intermediate language teachers. “Teachers cannot tell by looking at the exam how they should teaching reading or writing…(Wall and Alderson, 1996, p. 218)”. These programs can offer the necessary tools that teachers need to help their students achieve success with the tests in the intermediate language courses.

Typical in most academic institutions is the technique of measuring success by performance on tests. Understanding the teacher behaviors that encourage and help students to achieve success with tests, can assist administrators and support staff as they attempt to foster these behaviors with new teachers. Increased understanding of the teacher behaviors can help to encourage positive washback behaviors. At the higher
education level at a large institution such as the Ohio State University, suggestions for encouraging positive washback from the testing program in the classroom should be offered. In this environment, many of the teachers in the elementary and intermediate program are Graduate Teaching Assistants with limited experience. For some, their teaching repertoire is being developed and formulated, therefore they benefit from direct guidance.

According to the teacher observations, interviews, questionnaires, and the student questionnaires in this study, there are a number of things that can encourage teachers to demonstrate behaviors that not only reflect positive washback from tests, but also encourage skill development in the areas needed for the tests, and as described in the foreign language literature (speaking, listening, reading, and writing). These can include incorporating tasks that mimic real-life activities, using a variety of activities that focus on diverse skills, providing students with opportunities to use the target language to solve problems and gain new information, and using a variety of teaching techniques to meet the needs of all learners.

A characteristic of Spanish 104 in general brought to light by this study was limited instructional support, and instructional training provided to first time teachers, especially those with limited teaching experience. According to the study, this limited support appears to affect the efficacy of the teachers’ ability to effectively implement the Spanish 104 testing program. Although knowledge of the tests may have encouraged teachers to incorporate test like activities into their repertoires, without significant support and training in instructional techniques and effective pedagogy, teachers with
relatively little experience were teaching to the test demands within the scope of their limited abilities. They did not make a connection between skills and the assessments.

As Cheng (in Savignon, 2002) found, simply including information on a test does not guarantee that teachers will know how to teach the information to the students. “…the ways in which teachers perceive teaching and learning might change yet they are not likely to influence ways in which they teach. The change to a new exam has informed teachers about what they might do differently, but it has not shown them how to do it (p. 108). Inexperienced teachers may add these test-like activities to class time so that students will know better what to expect and be able to complete the test activities once they get to the test. Still, all of the teachers who participated in this study commented that they felt as though they needed further training to teach skills effectively at the intermediate level to encourage positive washback. With further training and support, new teachers to the course could expand their teaching techniques, lower their frustration, and theoretically do a better job of instructing the students in the area of skill building, thus better preparing them for the demands of the tests.

In this study, and in various educational settings, inexperienced teachers with limited teaching skills commonly exhibit behaviors that indicate disagreement with the format of a testing program, or they blame the test (‘it was confusing’, ‘unclear’, ‘too long’, ‘too hard’). Sometimes teachers fault the students (‘they didn’t study enough’, ‘they don’t get it’, ‘they shouldn’t be in this class’) when they perform poorly. With increased instructional training and direct support, teachers could positively modify their instructional practices to include more behaviors focused on test demands and skill development, thus promoting positive washback. Cheng (2002) purports that it is
important “…to recognize the need for conceptual change and teacher learning within the context of classroom practice. Any change would require sufficient time and resources for mentoring, peer coaching, intensive seminars, and the like” (p. 108).

An important detail observed in this study is that the teachers participating in the study responded positively to the focused attention that they received as participants. Although at first the teachers were intimidated having the researcher repeatedly in their classroom, they soon welcomed the attention. Teachers viewed the researcher as a resource, and a form of support. After two to three observations in each classroom, teachers began to ask the researcher questions about the course, ranging from procedural and syllabus questions, to questions about instructional methods, and theoretical application. Although it was not the purpose of the observations to critique the teaching, rather to observe and record, it was challenging to try to offer the feedback that each teacher searched for, while still trying to remain impartial.

It was difficult to not influence teacher behaviors based on what the researcher already knew about the course and the corresponding testing program. However, because these teachers saw the researcher as a resource, and because they valued the attention they received from being asked to participate in the study, they asked numerous questions, they were reflective of their own teaching, and they were aware of their students’ perceptions and reactions to their teaching. This implies that teachers teaching a new course for the first time want feedback and guidance for their teaching and the tests. Teachers appeared to want to be perceived as an effective teacher.

Whether this support comes in the form of focused professional development workshops, direct administrative support, or an organized mentoring program comprised
of experienced teachers, the newest teachers to an intermediate language program need
guidance. The data collected in this study, implies that teachers instructing within a new
curriculum for the first time, and using a new testing program markedly different than
what they have used in the past, want, and need support especially during the first quarter
of teaching the course. This support and guidance should provide instructional strategies
for teaching skills, especially in the area of reading and writing, and strategies for
meeting the demands of the corresponding testing program. Anne, Beth, Cathy, Diane,
and Elena indicated on their questionnaire that they believed that the students needed to
learn to read and write in order to perform on the tests. Each teacher seemed to realize
that skill development in these areas would help the students meet the demands of the
tests. However, they each struggled at some point with operationalizing this belief in
order to teach skills effectively within the context of the Spanish 104 course. There is a
need to train prospective teachers about effective ways to teach and develop these skills,
and then expand the skills beyond what is needed for the test. Then students can develop
skills to improve their language abilities and as a consequence, they can also be
successful on the tests.

Previous washback studies conducted by Wall and Alderson (1993), Alderson and
Hamp-Lyons (1996), Cheng (2002), and Qi (2004), have shown that affecting teacher
behaviors is challenging and requires an attentive focus on instruction. This study also
revealed a need for an attentive focus on instruction. Focusing on instruction can help to
provide teachers with the critical perspective needed for improvement, as well as the
impetus for change in instructional behaviors when needed. Without this type of
instructional focus, as was seen in this study and documented in other washback studies,
teachers will often continue to teach in the same manner that they are used to teaching, and will continue to emphasize what they believe are the important aspects of language learning, whether or not they are based on skills, and/or included on the tests.

5.4 Recommendations for Future Research

Although washback research has been conducted within language programs outside of the United States, to the researcher’s knowledge, this study is the only washback research conducted within the context of a Spanish foreign language program in the United States at the University level within a low stakes testing environment. In order to understand the teacher washback behaviors more thoroughly within foreign language classes, studies need to be conducted in classrooms at different language levels. Conducting the present study in an intermediate language classroom at the secondary level to determine if there are notable differences between the two levels could increase the generalizability of the present study. In addition, with the focus in American public schools on standardized testing, learning as much as possible about how tests affect teaching would provide important information to test writers, administrators, teachers, and students about washback. Further research in foreign languages other than Spanish can also help to determine if teacher washback behaviors in a foreign language class other than Spanish might demonstrate behaviors different than those described in this study.

The limited focus on students’ washback behaviors as they relate to teacher behaviors is another area that demands further research. As this study indicates, students’ behaviors appear to be affected by teachers’ instructional behaviors. In seeking to determine the degree of teacher washback, and the underlying teacher behaviors that
influence washback, it is important to understanding how this relates to students’ behaviors. Also as this study has shown, students have expectations with regard to required tests. Further understanding how these expectations impact student behaviors in the classroom is important for teachers to understand in order to plan for effective instruction.

Another avenue that has not been explored in washback research is the use of student test scores to measure washback. Although it would be challenging to measure the actual extent of washback by simply calculating the test means for students, the idea offers some interesting possibilities. A possible method for this is described within the context of the present study, but it would be theoretically applicable in other settings. The method outlined assumes a washback study taking place in the first course of an intermediate level program, following teachers who are teaching the course for the first time, and then again in the next quarter, or semester. The description presupposes a quarter system, however it could be easily conducted in a semester setting. Following teachers for an extended length of time (e.g. a year long course), could provide more substantial evidence than what the present study offered.

The first step in using test means to measure washback would be to calculate a student’s test mean from the course immediately preceding the intermediate level course. The previous course test mean would be used for the purpose of providing an historical profile for the individual student, in order to have more informed data when examining the intermediate level test mean. The researcher could presume that the average student would be expected to do at least as well in the intermediate course as in the previous course.
In order to use the students’ tests scores, it would be necessary to ensure that test scores are reliable and valid. Teachers would need to be trained in scoring the tests to ensure accuracy and inter-rater reliability. Once this has been established, the test scores can be used to calculate test means for students, and the teacher’s current class.

Next, a class test mean for the previous course and the intermediate course would be generated. This would allow a comparison between the class’ performance at the previous level with the class’ performance at the intermediate level. Two means (a student test mean, and a class test mean) would be generated for both groups of students. Upon conclusion of the data gathering process, the differences between the means from the two sections of each teacher would be compared. The comparisons would be made using weighted averages expressed as a percent, and a two sample t-test with significance set at $p = 0.05$. An increase in the test mean of the class from the second time teaching compared to the first time teaching for the same teacher, may provide the first evidence suggesting empirically observed washback.

A third set of means generated would be a mean test score for the all of a teachers’ students for the first time teaching and a mean test score for all of a teachers’ students from the second time teaching. To account for variability in class performance, the test mean for the previous course for both the first and second quarter classes would be used to adjust the test means as necessary. The mean class score for the first quarter would be expressed as a percent of the mean previous course score for the group to account for variation in student ability. The two means (first and second time teaching) would be compared to determine the amount of difference, and to investigate whether a difference may be suggestive of trends indicating washback. The first quarter grade
would be compared to the second quarter grade utilizing a two sample t-test with significance set at \( p = 0.05 \).

Although this alone cannot offer definitive evidence of washback because of the various factors that can influence test scores, it could certainly provide a perspective that has not been presented in any washback research thus far. Whereas students’ grades and performance are the most common ways to measure students’ understanding of material, and teachers’ effectiveness, it would be beneficial to try and quantify the possible effects of washback using actual student scores.

5.5 Conclusion

Washback has been and continues to be a challenging phenomenon to research and measure. Given the variables present in the classroom, this challenge persists in washback research. Although this study focused primarily on the teacher behaviors, the investigation into the students’ perceptions of washback, and the students’ behavior demonstrating washback reveals a need for further studies that focus on the students’ role in washback. This study showed that teachers do change their instructional habits after having taught a course only one time and these changes are a result of the influence of the testing program. Every teacher wants her students to perform well on tests as it is a reflection on her teaching.

Other important factors influencing changes in teachers’ behaviors reflect teachers’ beliefs, and the students’ desire to have accurate information about the testing program. Teacher beliefs influence teacher behaviors deeply no matter how the curriculum, course objectives, and the testing program work together. Students want information about the testing program because they want to be well prepared for the test
as a result of the classroom instruction. This study supports the notion that in order to affect the instruction that occurs in the classroom and create positive washback, direct guidance and training must be part of a teacher’s professional development. Even experienced teachers need professional development training in the areas that are new to them when teaching a course for the first time. As washback studies by Cheng, (2002), Watanabe (2004), and Shohamy, Donitsa-Schmidt, and Ferman (1996) have indicated, tests often provide teachers with new reasons for new activities; however, they do not necessarily show the teachers how to incorporate these activities into the curriculum.

Additionally, the tenets of past traditional language learning paradigms and their strong influence on present teacher behaviors as suggested in this study, must not be ignored. Whether the teachers in this study were once students in a foreign language program that was heavily influenced by grammar and structure, or whether they trained or worked in a program that influenced these areas in a previous teaching assignment, grammar appear to strongly influence teachers’ current instructional practices. This study shows that washback behaviors focusing on grammar and structure, are evident in instruction, regardless of the design of a language program, the format of the courses and the materials, and the demands of the testing program. Watanabe indicated similar teacher behaviors in his washback study (2004), as it seems that teachers will find a place in their repertoire for specific and direct grammar instruction. Focusing attention on this aspect, especially for first time teachers in an intermediate level, skill focused language class, can assist administrators and program directors as they help teachers to be more effective. More effective teachers mean happier and more successful students.
Tests will never be eliminated from educational institutions; therefore, it is best to embrace them and their power. Only then might they be used as a tool to bridge the gaps in instruction, and to train teachers not only to meet the demands of tests, but also meet the demands of their students, and affect their students’ achievement. Tests are only one component affecting change in the classroom; guidance and professional development are also necessary to the process. With focused guidance and professional development, this study implies that progress can be made toward creating positive washback from tests in the classroom both in terms of teacher behaviors and student behaviors. Tests can drive change, and if the intention is to make changes to foreign language instruction, not only at the intermediate level, but at all levels, it must be done consciously. Efforts must be taken to help teachers encourage positive washback and reduce negative washback.

In higher education (the context of this study), language education programs must focus on preparing teachers to teach well and teach effectively. Teachers at this level may need further training to balance their limited experience, and a testing program can provide the springboard to this training. For the Spanish 104 study, educational training and teaching experience influenced the documented changes in teacher behaviors that corresponded to the tests. The teacher who demonstrated the least notable changes in her instruction was the teacher with no training in Education, and limited teaching experience. With instructional support and guidance, teachers can focus instruction on skill development, and also to prepare the students for the tests. This research indicates that if skill development is encouraged in the class sessions, and then demanded on the tests, students presumably will view the skills as important, and consequently engage in washback behaviors. Additionally, teachers will also exhibit positive washback
behaviors, and thus set the example for the students that skills are the driving force in the classroom, as they happen to complement the tests. A positive consequence of washback behaviors that reflect skills can be successful test performance. This, in turn, provides satisfaction to students, teachers, and program administrators. Skills are the focus in the intermediate level Spanish 104 language program, and the corresponding testing program. Therefore, more needs to be done to assure that this message is understood clearly, not because there is a skills based test, but because teachers and students need to be informed, and then equipped to meet the demands and expectations.
BIBLIOGRAPHY


APPENDIX A

CLASS OBSERVATION INSTRUMENT

Time

T (teacher)

S (students)

Date

Instructor
APPENDIX B

TEACHER INSTRUMENTS

Spanish 104 washback project
First teacher interview

Name: _____________________  Age:______
Degree held:_______________  Training received: ____________________
No. of years teaching Spanish: __________
Date of interview: _____________________

Time began: ______________  Time concluded: ______________

Introductory statement: Thank you for agreeing to work with me on this project. The purpose of this project is to determine the washback effects of testing on teaching in Spanish 104. I will ask you some questions about your expectations for teaching Spanish 104, and how you plan to prepare your students for the testing program in this level. What you tell me will be used for this research project only, and is completely unrelated to any department evaluations. With your permission, I will tape record all of our interviews to ensure accuracy in recording your responses. Please answer the questions as completely and as honestly as you can.

1. What is your teaching plan for the course?

2. Describe your perception of the transition between the elementary level courses (Spanish 101, 102.66/102.01, 103.66/103.01) and Spanish 104.

3. What types of activities do you envision will encompass daily classroom activity?

4. What is your understanding of the link between testing and assessment?
5. How do you plan to prepare your students for the tests?

6. How will you prepare them for the specific sections?
   writing:
   reading:
   listening:

7. How will classroom activities reflect testing expectations?

8. Is there anything else about the testing program or about Spanish 104 in general that you would like to comment on or share with me?
Spanish 104 washback project
Second teacher interview

Name: _____________________
Time began: ______________
Time concluded: ______________

Introductory statement: Thank you for continuing to work with me on this project. As you know, the purpose of this project has been, and continues to be, to determine the washback effects of testing on teaching in Spanish 104. After having given two tests, please comment on how the testing program has affected your teaching plan. Again, anything you tell me will be used for this research project only, and is in no way related to departmental evaluations. Please answer the questions as completely and as honestly as you can.

1. What adjustments to your teaching have you made since the first examination?

2. How have the students responded to the examination process?

3. What activities have you eliminated from class sessions that you had initially planned on incorporating?

4. How has your knowledge of the testing process affected the way you plan for and conduct class instruction?

5. In what areas do you feel that you need to improve instruction in order to help students better prepare for the demands of the tests?

6. Additional comments that you wish to share about the effect of testing on your teaching in Spanish 104:
Teacher’s perceptions of exams in Spanish 104

Please answer the following questions by grading on a 5 point Likert scale format where
1=Strongly disagree
2=Disagree
3=Neutral
4=Agree
5=Strongly Agree

Write 1,2,3,4, or 5 in the brackets.

1. What kind of extra work or pressure if any do you think the Spanish 104 tests have put on you in your teaching?
[ ] Following a new syllabus
[ ] Doing more lesson preparation
[ ] Preparing more materials for students
[ ] Revising the existing materials that you have
[ ] Employing new teaching methods
[ ] Setting up new teaching objectives
[ ] Meeting new challenges in teaching
[ ] Organizing more focused activities that reflect exam activities

2. What are the most significant changes you have made in your teaching in the context of the different exam format for the 104 tests?
[ ] Teaching according to the test format
[ ] Adopting new teaching methods
[ ] Using a more communicative approach in teaching
[ ] Emphasizing role play and group discussions
[ ] Emphasizing listening skills
[ ] Emphasizing the integration of skills
[ ] Employing more real life language tasks
[ ] Encouraging more student participation in class

3. What do you find the most difficult or challenging aspects of teaching Spanish 104?
[ ] Students’ current level of Spanish
[ ] Class size
[ ] Inadequate teaching resources
[ ] Learning environment
[ ] Limited teaching and learning aids and facilities
[ ] Overly demanding workload
[ ] Inadequate time for students to practice Spanish outside of the classroom
4. What are the learning strategies you would recommend to your students in the context of the transition between the elementary language program and the intermediate language program?

- Learn to take better notes
- Expose themselves to various Spanish realia outside of the classroom
- Express their opinions more frequently in class
- Put more emphasis on speaking
- Put more emphasis on listening
- Put more emphasis on reading
- Put more emphasis on writing
- Learn to initiate questions
- Be more active in class participation
- Create opportunities to use Spanish in daily life
- Transition from passive learning to active learning
- Communicate more in Spanish

5. What types of activities do you believe are most effective in language learning?

- Task-oriented activities
- Language games
- Role play and group discussion
- Authentic materials
- Training in basic language knowledge
- Extracurricular activities

6. What you believe to be the major aims for studying and learning Spanish at the Ohio State University?

- To pursue further studies
- To pass examinations
- To meet requirements
- To obtain jobs

7. In what ways would you like to motivate your students to learn Spanish?

- Do more mock exam like activities
- Use more authentic materials
- Organize real life language activities
- Use more interesting language games
- Give students more encouragement to learn
- Create a positive attitude toward language learning
- Provide students with effective language learning strategies
8. The factors that most influence your teaching are:
   [ ] Professional training
   [ ] Academic seminars or workshops
   [ ] Teaching experience and beliefs
   [ ] Teaching syllabus
   [ ] Past experience as a language learner
   [ ] The need to obtain satisfaction in teaching
   [ ] Textbook
   [ ] Exams
   [ ] Learners’ expectations
   [ ] Peers’ expectations
   [ ] Supervisors’ expectations
   [ ] Social expectations
Spanish 104 washback project
Third teacher interview

Name: ____________________
Time began: ______________
Time concluded: ________________

Introductory statement: Thank you for your assistance on this project. As you know, the purpose of this project has been, and continues to be, to determine the washback effects of testing on teaching in Spanish 104. After having taught Spanish 104 for a quarter, please comment on how the testing program affected your teaching and vice versa. Again, anything you tell me will be used for this research project only, and is in no way related to departmental evaluations. Please answer the questions as completely and as honestly as you can.

1. How did your teaching plan for the course change as a result of the administration of the first test?

2. Did your perception of the transition between the elementary level courses (Spanish 101, 102.66/102.01, 103.66/103.01) and Spanish 104 change during the period of time from the commencement of the quarter to the conclusion? How?

3. Having administered the testing program for a quarter, what do you believe to be the best way (s) help the students prepare for the tests?

4. How did you help them prepare for the specific sections of each test?

   writing:
   reading:
   listening:

5. How can the articulation between the elementary levels and the intermediate 104 level be improved?
6. How has your understanding of the link between testing and assessment changed during the quarter?

7. Do you feel that the expected classroom activities as outlined in the course and departmental objectives accurately reflect testing expectations?

8. Further comments that you wish to share about the effect of testing on your teaching in Spanish 104:
Spanish 104 washback project
Second quarter teacher interview

Name: _____________________
Time began: ____________ Time concluded: ____________

Introductory statement: Thank you for your continued support on this project. As you know, the purpose of this project has been, and continues to be, to determine the washback effects of testing on teaching in Spanish 104. After having taught Spanish 104 for one quarter, please comment on how knowledge of the testing program has affected your teaching plan. Again, anything you tell me will be used for this research project only, and is in no way related to departmental evaluations. With your permission, I will tape record our interview to ensure accuracy in recording your responses. Please answer the questions as completely and as honestly as you can.

1. What adjustments to your teaching have you made this quarter since the administration of the first test during the autumn quarter?

2. How did the way you prepared students for the first test this quarter differ from the autumn quarter when you had limited knowledge of the testing program?

3. What activities have you incorporated into your class sessions that specifically focus on aspects that the tests address?

4. How has your knowledge of the testing process affected the way you plan for and conduct class instruction this quarter as opposed to autumn quarter?

5. In what areas do you feel that you still need to improve instruction in order to help students better prepare for the demands of the tests?

6. Additional comments that you wish to share about differences in your teaching that you have noticed between the autumn and winter quarters in Spanish 104:
Student surveys: Chapter 1, 2, and 3 Tests

**PLEASE DO NOT PUT YOUR NAME ANYWHERE ON THIS SURVEY**

Please answer the following questions honestly by circling your response, or providing the information as indicated. The results of this survey will not be shared with your instructor during the course of your class.

1. How did you study for the following sections of the test?

   a. Listening comprehension: in-class on your own both didn’t study
   b. Reading comprehension: in-class on your own both didn’t study
   c. Questions about readings: in-class on your own both didn’t study
   d. Structure and grammar: in-class on your own both didn’t study

2. How much time did you devote to studying for the following sections of the test?

<table>
<thead>
<tr>
<th></th>
<th>Hours</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questions about readings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure and grammar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. How many hours per week were dedicated to preparing for the upcoming tests?

   0 15 minutes 30 minutes 1 hour 1.5 hours 2 hours 3 hours
Spanish 104 washback project
Student questionnaire for Spanish 104 final exam

Student survey for Spanish 104
This survey is being conducted for research purposes regarding the influence of testing on instruction in Spanish 104. The answers provided on this questionnaire are NOT related to your grade for the course, nor will the researcher be informed of your personal identity. The researcher asks permission to access your test grades from Spanish 103 and Spanish 104. The purpose is to compute a class test average for SP 103 in order to compare it to the class test average for SP 104. Please check the appropriate box, and provide the corresponding information as necessary. The researcher will NOT be able to match your personal identity to the grade, and the information will be used for research purposes only. Please select one.

All information provided will be held confidential. Please do not put your name anywhere on the survey. Thank you for your time and your honest answers.

Please check the appropriate box. 

<table>
<thead>
<tr>
<th>Permission granted to access Spanish 103 and 104 test average.</th>
<th></th>
<th>Grades will be anonymously averaged. Please continue with the questionnaire.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Permission not granted to access Spanish 103 and 104 test average.</th>
<th>Write the last 4 digits of your SS# to eliminate your information from the class average. Please continue with the questionnaire.</th>
</tr>
</thead>
</table>

Please fill in the top portion of this survey prior to answering the questions.

Male ____  Female____  Major @ OSU:______________________________

Year in school: ____________  Age: ____________

Spanish classes taken at OSU prior to 104: ____________________________

Quarter(s) taken: ______________________

Anticipated grade in Spanish 104: ________
For both Part I and Part II, please answer the following questions truthfully, using the following Likert type scale where

1=Not at all
2=Not significantly
3=Somewhat/Slighly
4=Tremendously

Please circle the number that best represents your answer for the question.

---

Part I.

1. The class activities in Spanish 104 were different than those of my previous Spanish classes.
   1 2 3 4

2. The class activities helped me to prepare appropriately for the tests.
   1 2 3 4

3. I studied for the test by engaging in activities similar to class activities.
   1 2 3 4

4. I studied for the test with classmates.
   1 2 3 4

5. I engaged in the following activities to study for the test:
   - reading 1 2 3 4
   - writing 1 2 3 4
   - speaking 1 2 3 4
   - listening 1 2 3 4
   - group work 1 2 3 4

6. I frequently seek opportunities to use Spanish outside of class.
   1 2 3 4

7. The teacher/instructor clearly communicated the expectations for the test.
   1 2 3 4
Part II.

8. When you studied for the **listening comprehension** portion of the test, you improved your language skills as follows:

   - **listening**: 1 2 3 4
   - **reading**: 1 2 3 4
   - **writing**: 1 2 3 4
   - **speaking**: 1 2 3 4

2. When you studied for the **reading comprehension** portion of the test, you improved your language skills as follows:

   - **listening**: 1 2 3 4
   - **reading**: 1 2 3 4
   - **writing**: 1 2 3 4
   - **speaking**: 1 2 3 4

3. When you studied for the **questions about the readings from the book** portion of the test, you improved your language skills as follows:

   - **listening**: 1 2 3 4
   - **reading**: 1 2 3 4
   - **writing**: 1 2 3 4
   - **speaking**: 1 2 3 4

4. When you studied for the **grammar and structure** portion of the test, you improved your language skills as follows:

   - **listening**: 1 2 3 4
   - **reading**: 1 2 3 4
   - **writing**: 1 2 3 4
   - **speaking**: 1 2 3 4
APPENDIX D

INSTRUCTIONAL TIME DISTRIBUTION (TEACHER CLASSROOM OBSERVATIONS AUTUMN 2004; WINTER 2005)

<table>
<thead>
<tr>
<th>Teacher: Anne</th>
<th>Quarter: Autumn</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Time spent (Minutes)</th>
<th>Instructional time (Percentage)</th>
<th>Activity/instruction</th>
</tr>
</thead>
</table>
| **Lesson 1**  
Test impacting | | |
| 2 | 4 | mentions reading/grammar on test |
| 13 | 27 | oral vocab activities |
| 9 | 18.7 | grammar (writing) |
| 6 | 12.5 | *a* reading- no specific instruction |
| 4 | 8.3 | listening comprehension |
| 4 | 8.3 | oral grammar activity (gustar) |
| 8 | 16.6 | administrative (re: class) |
| **Lesson 2**  
Test impacting | | |
| 16 | 34 | song for grammar (imperfect tense) |
| 19 | 40 | during reading activities Sak-Nicté |
| 12 | 25 | post-reading activities |
| **Lesson 3**  
Non testing impacting | | |
<p>| 15 | 31 | correcting HW (listening comprehension; Diario) |
| 22 | 45.8 | grammar explanation/practice (preterite) |
| 11 | 22.9 | grammar practice (oral- preterite) |</p>
<table>
<thead>
<tr>
<th>Lesson 4</th>
<th>Test impacting</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>47.9</td>
</tr>
<tr>
<td></td>
<td>pre-reading activities, initial story discussion</td>
</tr>
<tr>
<td>16</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>during reading activities, questions</td>
</tr>
<tr>
<td>9</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>post-reading, (act out and literary terms)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 5</th>
<th>Non test impacting</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>grammar explanation/practice (I/O D/O pronouns)</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>listening comprehension activities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 6</th>
<th>Test impacting</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>85.4</td>
</tr>
<tr>
<td></td>
<td>grammar instruction/practice (subjunctive)</td>
</tr>
<tr>
<td>7</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>group activity (subjunctive) from text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 7</th>
<th>Test impacting</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>grammar instruction/practice (imperfect subjunctive)</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>story activity with imperfect subjunctive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8</th>
<th>Non test impacting</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>grammar activity from text (“if” clauses)</td>
</tr>
<tr>
<td>6</td>
<td>13.3</td>
</tr>
<tr>
<td></td>
<td>grammar activity (“if clauses) oral</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>grammar activity (“if” clauses) oral and written “What if…” game</td>
</tr>
</tbody>
</table>

Table A.1 Instructional behaviors for Anne’s observed classes during Autumn 2004
<table>
<thead>
<tr>
<th>Teacher: Beth</th>
<th>Quarter: Autumn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time spent (Minutes)</strong></td>
<td><strong>Instructional time (Percentage)</strong></td>
</tr>
<tr>
<td><strong>Lesson 1</strong> Test impacting</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td><strong>Lesson 2</strong> Test impacting</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>12.5</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
</tr>
<tr>
<td><strong>Lesson 3</strong> Non test impacting</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17.7</td>
</tr>
<tr>
<td>11</td>
<td>24.4</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td><strong>Lesson 4</strong> Test impacting</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>41.6</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Lesson 5 Non test impacting</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--</td>
</tr>
<tr>
<td>11</td>
<td>22.9</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>24</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 6 Test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>33.3</td>
<td>grammar activities (subjunctive) text/teacher</td>
</tr>
<tr>
<td>28</td>
<td>58.3</td>
<td>grammar activity (Dear Abby) teacher</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>passes back papers/offers comments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 7 Test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>60</td>
<td>presentations</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>grammar activity (conditional) oral</td>
</tr>
<tr>
<td>6</td>
<td>12.5</td>
<td>grammar activity (conditional) text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8 Non test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>37.5</td>
<td>grammar activity; text</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
<td>grammar activity; text</td>
</tr>
<tr>
<td>15</td>
<td>31.3</td>
<td>grammar activity; overhead</td>
</tr>
<tr>
<td>7</td>
<td>14.58</td>
<td>passes back and goes over second exam</td>
</tr>
</tbody>
</table>

Table A.2 Instructional behaviors for Beth’s observed classes during Autumn 2004
<table>
<thead>
<tr>
<th>Teacher: Cathy</th>
<th>Quarter: Autumn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time spent</strong></td>
<td><strong>Instructional time</strong></td>
</tr>
<tr>
<td><strong>(Minutes)</strong></td>
<td><strong>(Percentage)</strong></td>
</tr>
<tr>
<td><strong>Lesson 1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Test impacting</strong></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17.7</td>
</tr>
<tr>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>15.5</td>
</tr>
<tr>
<td>16</td>
<td>35.5</td>
</tr>
<tr>
<td><strong>Lesson 2</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Test impacting</strong></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>33.3</td>
</tr>
<tr>
<td>17</td>
<td>35.4</td>
</tr>
<tr>
<td>15</td>
<td>31.25</td>
</tr>
<tr>
<td><strong>Lesson 3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Non test impacting</strong></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>18.7</td>
</tr>
<tr>
<td>21</td>
<td>43.75</td>
</tr>
<tr>
<td>2-3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
</tr>
<tr>
<td>5</td>
<td>10.4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Lesson 4</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Test impacting</strong></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>14.5</td>
</tr>
<tr>
<td>11</td>
<td>22.9</td>
</tr>
<tr>
<td>15</td>
<td>31.25</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Lesson 5</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Non test impacting</strong></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>41.6</td>
</tr>
<tr>
<td>4</td>
<td>8.3</td>
</tr>
<tr>
<td>Lesson 6</td>
<td>Test impacting</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>10.4</td>
</tr>
<tr>
<td>19</td>
<td>39.5</td>
</tr>
<tr>
<td>9</td>
<td>18.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 7</th>
<th>Test impacting</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>25</td>
<td>oral practice- chapter theme- grammar based</td>
</tr>
<tr>
<td>10</td>
<td>20.8</td>
<td>grammar- oral book activities</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>explanation of grammar rules</td>
</tr>
<tr>
<td>7</td>
<td>14.58</td>
<td>grammar practice- activities from book</td>
</tr>
<tr>
<td>16</td>
<td>33.3</td>
<td>review activities (theme related)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8</th>
<th>Non test impacting</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>14.58</td>
<td>grammar explanation on board</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>grammar practice</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>transition to new activity (set up)</td>
</tr>
<tr>
<td>14</td>
<td>29</td>
<td>song activity stressing grammar</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>grammar- ‘si’ clauses</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
<td>grammar- imperfect subjunctive review and practice</td>
</tr>
</tbody>
</table>

Table A.3 Instructional behaviors for Cathy’s observed classes during Autumn 2004
<table>
<thead>
<tr>
<th>Teacher:</th>
<th>Quarter:</th>
<th>Time spent</th>
<th>Instructional time</th>
<th>Activity/instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diane</td>
<td>Autumn</td>
<td>(Minutes)</td>
<td>(Percentage)</td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test impacting</td>
<td>19</td>
<td>39.5</td>
<td>music vocab question/ activities on OH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12.5</td>
<td>OH questions/ pre-reading activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>29.16</td>
<td>Gloria Estefan story and questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>22.9</td>
<td>music in general; oral work OH</td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test impacting</td>
<td>15</td>
<td>31.25</td>
<td>warm-up grammar; corrections of writing (related to required story reading)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>56.25</td>
<td>vocabulary activity- writing from stories</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12.5</td>
<td>OH grammar- questions on preterite/imperfect</td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non test impacting</td>
<td>6</td>
<td>12.5</td>
<td>set up/questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>62.5</td>
<td>listening activities (Diario)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>video set up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>22.9</td>
<td>video activities</td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test impacting</td>
<td>3</td>
<td>6.5</td>
<td>set up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>39</td>
<td>pre-reading activities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>39</td>
<td>analysis of story/organization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>13</td>
<td>story <em>(Preguntas de Orientación)</em></td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non test impacting</td>
<td>6</td>
<td>12.5</td>
<td>set up/organization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>87.5</td>
<td>listening comprehension activities (OH)</td>
<td></td>
</tr>
<tr>
<td>Lesson 6</td>
<td>Test impacting</td>
<td>2</td>
<td>4</td>
<td>set up/organization</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>60</td>
<td>vocab activities- group work and student participation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>35</td>
<td>grammar- subjunctive instruction OH</td>
</tr>
<tr>
<td>Lesson 7</td>
<td>Test impacting</td>
<td>5</td>
<td>10.4</td>
<td>set up/ organization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>31.2</td>
<td>vocab activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>28</td>
<td>58</td>
<td>required reading and post reading activities</td>
</tr>
<tr>
<td>Lesson 8</td>
<td>Non test impacting</td>
<td>13</td>
<td>27</td>
<td>grammar review- conditional tense</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35</td>
<td>72.9</td>
<td>grammar review- present and imperfect subjunctive OH activities</td>
</tr>
</tbody>
</table>

Table A.4 Instructional behaviors for Diane’s observed classes during Autumn 2004
<table>
<thead>
<tr>
<th>Lesson</th>
<th>Non test impacting</th>
<th>Test impacting</th>
<th>(Minutes)</th>
<th>Instructional time (Percentage)</th>
<th>Activity/instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1</td>
<td>Test impacting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>50</td>
<td>reading/discussion and activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>29</td>
<td>interview activity with students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>20.8</td>
<td>poem reading- analysis and meaning</td>
</tr>
<tr>
<td>Lesson 2</td>
<td>Test impacting</td>
<td></td>
<td>3</td>
<td>6</td>
<td>set up video machines</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>45</td>
<td>93.7</td>
<td>videos and activites for chapters 1 &amp; 2</td>
</tr>
<tr>
<td>Lesson 3</td>
<td>Non test impacting</td>
<td></td>
<td>2</td>
<td>4</td>
<td>overview about test (on board)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>16.6</td>
<td>pre-reading activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>47.9</td>
<td>required reading/discussion, comprehension check and activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>31</td>
<td>post-reading activity</td>
</tr>
<tr>
<td>Lesson 4</td>
<td>Test impacting</td>
<td>(5 min. late)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>81</td>
<td>reading activities/ comprehension</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>18.6</td>
<td>analysis of story; theme</td>
</tr>
<tr>
<td>Lesson 5</td>
<td>Non test impacting</td>
<td>(5 min. late)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>passing out materials</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>42</td>
<td>97.6</td>
<td>grammar instruction and review; practice-indirect/direct object pronouns</td>
</tr>
<tr>
<td>Lesson 6</td>
<td>Test impacting</td>
<td></td>
<td>3</td>
<td>6</td>
<td>explains “bilingue- traductor- interprete”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>grammar explanation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>42</td>
<td>91</td>
<td>student presentations</td>
</tr>
<tr>
<td>Lesson 7</td>
<td>Test impacting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>story discussion and activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesson 8</td>
<td>Non test impacting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>16.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>grammar- preterite/imperfect explanation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>grammar “quiz”- students use books, correct with teacher and move on to next section</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.5 Instructional behaviors for Elena’s observed classes during Autumn 2004
<table>
<thead>
<tr>
<th>Teacher: Anne</th>
<th>Quarter: Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time spent (Minutes)</strong></td>
<td><strong>Instructional time (Percentage)</strong></td>
</tr>
<tr>
<td><strong>Lesson 1</strong></td>
<td><strong>Lesson 2</strong></td>
</tr>
<tr>
<td>Test impacting</td>
<td>Test impacting</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>49</td>
</tr>
<tr>
<td>10</td>
<td>18.7</td>
</tr>
<tr>
<td>13</td>
<td>81.12</td>
</tr>
<tr>
<td>set up/introduction</td>
<td>explains test (format/content)</td>
</tr>
<tr>
<td>story and activity</td>
<td></td>
</tr>
<tr>
<td>grammar</td>
<td></td>
</tr>
<tr>
<td>Lesson 7</td>
<td>Test impacting</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>17</td>
<td>37.7</td>
</tr>
<tr>
<td>7</td>
<td>15.5</td>
</tr>
<tr>
<td>21</td>
<td>46.6</td>
</tr>
<tr>
<td></td>
<td>vocabulary activity (theme of chapter)</td>
</tr>
<tr>
<td></td>
<td>grammar introduction</td>
</tr>
<tr>
<td></td>
<td>contextualized grammar activities within theme of chapter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8</th>
<th>Non test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>62.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>37.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>video (background, viewing activity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>grammar- instruction/activity; conditional/ imperfect subjunctive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.6 Instructional behaviors for Anne’s observed classes during Winter 2005
<table>
<thead>
<tr>
<th>Teacher: Beth</th>
<th>Quarter: Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time spent (Minutes)</strong></td>
<td><strong>Instructional time (Percentage)</strong></td>
</tr>
<tr>
<td><strong>Lesson 1</strong>&lt;br&gt;Test impacting</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>33.3</td>
</tr>
<tr>
<td>31</td>
<td>64.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Lesson 2</strong>&lt;br&gt;Test impacting</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>10.4</td>
</tr>
<tr>
<td>15</td>
<td>31.2</td>
</tr>
<tr>
<td><strong>Lesson 3</strong>&lt;br&gt;Non test impacting</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>28.8</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td><strong>Lesson 4</strong>&lt;br&gt;Test impacting</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>10.4</td>
</tr>
<tr>
<td>36</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
</tr>
<tr>
<td><strong>Lesson 5</strong>&lt;br&gt;Non test impacting</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6.9</td>
</tr>
<tr>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>9</td>
<td>20.9</td>
</tr>
<tr>
<td>15</td>
<td>34.8</td>
</tr>
<tr>
<td>13</td>
<td>30.2</td>
</tr>
<tr>
<td>2</td>
<td>4.6</td>
</tr>
<tr>
<td>Lesson 6</td>
<td>Test impacting</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Lesson 7</td>
<td>Test impacting</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Lesson 8</td>
<td>Non test impacting</td>
</tr>
<tr>
<td></td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
</tbody>
</table>

Table A.7 Instructional behaviors for Beth’s observed classes during Winter 2005
<table>
<thead>
<tr>
<th>Lesson</th>
<th>Test impacting</th>
<th>Time spent (Minutes)</th>
<th>Instructional time (Percentage)</th>
<th>Activity/instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1</td>
<td>Test impacting</td>
<td>7</td>
<td>11.8</td>
<td>introduction to materials, etc. (second class)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>22</td>
<td>grammar- introduction and practice (PE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39</td>
<td>66</td>
<td>required reading and activity</td>
</tr>
<tr>
<td>Lesson 2</td>
<td>Test impacting</td>
<td>11</td>
<td>18.9</td>
<td>grammar- oral practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47</td>
<td>81</td>
<td>required reading and activities (specific to test related activities; comprehensive discussion, etc.)</td>
</tr>
<tr>
<td>Lesson 3</td>
<td>Non test impacting</td>
<td>35</td>
<td>74.4</td>
<td>grammar- practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>12.76</td>
<td>test information (format and content)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>12.76</td>
<td>grammar review- white boards</td>
</tr>
<tr>
<td>Lesson 4</td>
<td>Test impacting</td>
<td>38</td>
<td>67.8</td>
<td>grammar- instruction and practice indirect/direct objects (text)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>32</td>
<td>grammar- review and practice comparisons</td>
</tr>
<tr>
<td>Lesson 5</td>
<td>Non test impacting</td>
<td>42</td>
<td>75</td>
<td>required reading and activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>25</td>
<td>discussion of presentations</td>
</tr>
<tr>
<td>Lesson 6</td>
<td>Test impacting</td>
<td>21</td>
<td>36.8</td>
<td>pre-reading activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>53.5</td>
<td>required reading and post reading activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>10.7</td>
<td>grammar- practice subjunctive</td>
</tr>
<tr>
<td>Lesson 7 Test impacting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>administrative tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>46.9</td>
<td>pre-reading/vocabulary activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>37.8</td>
<td>reading and summarizing activity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8 Non test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>80</td>
<td>post reading activity</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>grammar- extracted from activities (thematic)</td>
</tr>
</tbody>
</table>

Table A.8 Instructional behaviors for Cathy’s observed classes during Winter 2005
<table>
<thead>
<tr>
<th>Teacher:</th>
<th>Quarter:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diane</td>
<td>Winter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time spent (Minutes)</th>
<th>Instructional time (Percentage)</th>
<th>Activity/instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson 1</strong> Test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>29</td>
<td>pre-reading activity</td>
</tr>
<tr>
<td>25</td>
<td>52</td>
<td>required reading</td>
</tr>
<tr>
<td>9</td>
<td>18.75</td>
<td>post reading</td>
</tr>
<tr>
<td><strong>Lesson 2</strong> Test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>67.4</td>
<td>mini quiz- reading and post reading</td>
</tr>
<tr>
<td>14</td>
<td>32.5</td>
<td>grammar- instruction and practice</td>
</tr>
<tr>
<td><strong>Lesson 3</strong> Non test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>39.5</td>
<td>reading- review and content</td>
</tr>
<tr>
<td>20</td>
<td>46.5</td>
<td>grammar- review and activities</td>
</tr>
<tr>
<td>5</td>
<td>10.86</td>
<td>acting out story and read aloud</td>
</tr>
<tr>
<td>4</td>
<td>8.3</td>
<td>information about tests (content and format) OH</td>
</tr>
<tr>
<td><strong>Lesson 4</strong> Test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>83.3</td>
<td>pre-reading activities</td>
</tr>
<tr>
<td>8</td>
<td>18.6</td>
<td>reading</td>
</tr>
<tr>
<td><strong>Lesson 5</strong> Non test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>45.6</td>
<td>grammar-practice activity superlatives</td>
</tr>
<tr>
<td>15</td>
<td>32.6</td>
<td>grammar- practice activity double object pronouns</td>
</tr>
<tr>
<td>10</td>
<td>21.7</td>
<td>project discussion/ explanation of evaluation</td>
</tr>
<tr>
<td><strong>Lesson 6</strong> Test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>70.2</td>
<td>pre-reading activities, vocab</td>
</tr>
<tr>
<td>6</td>
<td>12.76</td>
<td>reading</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>post reading activity</td>
</tr>
</tbody>
</table>
### Lesson 7
Test impacting

<table>
<thead>
<tr>
<th>28</th>
<th>54</th>
<th>pre-reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>45.8</td>
<td>reading and post reading activities</td>
</tr>
</tbody>
</table>

### Lesson 8
Non test impacting

<table>
<thead>
<tr>
<th>14</th>
<th>29</th>
<th>grammar- instruction and practice conditional activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>70.8</td>
<td>grammar- instruction and practice imperfect subjunctive activities</td>
</tr>
</tbody>
</table>

Table A.9 Instructional behaviors for Diane’s observed class during Winter 2005
<table>
<thead>
<tr>
<th>Lesson</th>
<th>Type</th>
<th>Activity/instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson 1</strong></td>
<td>Test impacting</td>
<td>Reading</td>
</tr>
<tr>
<td>46</td>
<td>100</td>
<td>reading</td>
</tr>
<tr>
<td><strong>Lesson 2</strong></td>
<td>Test impacting</td>
<td>Pre-reading</td>
</tr>
<tr>
<td>8</td>
<td>16.6</td>
<td>pre-reading</td>
</tr>
<tr>
<td>20</td>
<td>41.6</td>
<td>reading and discussion</td>
</tr>
<tr>
<td>9</td>
<td>18.75</td>
<td>post reading</td>
</tr>
<tr>
<td>11</td>
<td>22.9</td>
<td>listening comprehension</td>
</tr>
<tr>
<td><strong>Lesson 3</strong></td>
<td>Non-test impacting</td>
<td>Introduction/warm up</td>
</tr>
<tr>
<td>6</td>
<td>13.3</td>
<td>introduction/warm up</td>
</tr>
<tr>
<td>11</td>
<td>24.4</td>
<td>grammar- review sheet</td>
</tr>
<tr>
<td>15</td>
<td>33.3</td>
<td>reading summaries</td>
</tr>
<tr>
<td>5</td>
<td>11.1</td>
<td>listening comprehension</td>
</tr>
<tr>
<td>8</td>
<td>17.7</td>
<td>grammar- practice (book) preterite/imperfect</td>
</tr>
<tr>
<td><strong>Lesson 4</strong></td>
<td>Test impacting</td>
<td>Composition information and preparation</td>
</tr>
<tr>
<td>25</td>
<td>54</td>
<td>composition information and preparation</td>
</tr>
<tr>
<td>21</td>
<td>45.65</td>
<td>teacher instruction about writing</td>
</tr>
<tr>
<td><strong>Lesson 5</strong></td>
<td>Non-test impacting</td>
<td>Pre-reading</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>pre-reading</td>
</tr>
<tr>
<td>13</td>
<td>29.5</td>
<td>article and discussion</td>
</tr>
<tr>
<td>11</td>
<td>25</td>
<td>listening comprehension</td>
</tr>
<tr>
<td><strong>Lesson 6</strong></td>
<td>Test impacting</td>
<td>Reading activity (student work- although teacher focused entirely on reading, she never brought class back together, however did collect the work)</td>
</tr>
<tr>
<td>48</td>
<td>100</td>
<td>reading activity (student work- although teacher focused entirely on reading, she never brought class back together, however did collect the work)</td>
</tr>
<tr>
<td>Lesson 7 Test impacting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>42.5</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>51</td>
</tr>
<tr>
<td>warm up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading and discussion (thorough and supplemental)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading summary activity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson 8 Non test impacting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12.5</td>
</tr>
<tr>
<td>late to class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>passing out paper/information about projects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grammar- review and practice worksheet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grammar- writing contextualized to theme of chapter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.10 Instructional behaviors for Elena’s observed classes during Winter 2005