STUDENTS’ AND TEACHERS’ PERCEPTIONS OF EFFECTIVE TEACHING IN THE FOREIGN LANGUAGE CLASSROOM: A COMPARISON OF IDEALS AND RATINGS

by

Alan Victor Brown

A Dissertation Submitted to the Faculty of the GRADUATE INTERDISCIPLINARY PROGRAM IN SECOND LANGUAGE ACQUISITION AND TEACHING

In Partial Fulfillment of the Requirements For the Degree of DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2006
THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Alan Victor Brown entitled “Students’ and Teachers’ Perceptions of Effective Teaching in the Foreign Language Classroom: A Comparison of Ideals and Ratings” and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_________________________________________ Date: May 25, 2006
Dr. Renate Schulz

_________________________________________ Date: May 25, 2006
Dr. Beatrice Dupuy

_________________________________________ Date: May 25, 2006
Dr. Mary Wildner-Bassett

_________________________________________ Date: May 25, 2006
Dr. Lawrence Aleamoni

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_________________________________________ Date: May 25, 2006
Dissertation Director: Renate Schulz
STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ALAN V. BROWN
ACKNOWLEDGEMENTS

This dissertation is the result of the efforts and expertise of many different people who deserve my most sincere gratitude. The faculty who served on my committee and graciously offered their time and talents deserve special thanks as they helped polish all aspects of this dissertation, from the original idea to the final draft. They are Dr. Renate Schulz, Dr. Beatrice Dupuy, Dr. H.D. Adamson, Dr. Mary Wildner-Bassett, and Dr. Lawrence Aleamoni. My director, Dr. Renate Schulz, demonstrated a level of patience only my persistent requests for feedback could evoke.

Clearly this dissertation could not have been brought to fruition without those teachers and students who allowed me to interrupt their classes and then willingly participated in my study—all for the purposes of ‘research.’ It is to them that I owe the greatest expression of appreciation for without them I would have nothing to report on.

Rolando Garcia de Alba, Peter In-Albon, and Mark Borgstrom from various departments across campus graciously invested their time and expertise into this project.

Without a doubt, the one individual who contributed the most to the completion of this dissertation was my dearest wife. Andrea not only provided much needed moral support but also spent hours and hours helping me during the data collection phase by engaging in the most tedious chores such as checking timing marks on scantron sheets, stapling questionnaires, preparing packets for administration, rushing to administer questionnaires across campus, and finding someone to watch our lovely daughter—Gracie. It is thanks to Andrea that I was able to finish this dissertation, and more importantly, understand that dissertations are not nearly as important as family.
DEDICATION

I dedicate this dissertation to my wonderful mother and father who have taught me more about hard work, perseverance, courage, and love than a hundred dissertations ever could. I can only thank God that I had the opportunity to be tutored so closely about life from people as tenacious and loving as my parents.
TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................. 10
LIST OF TABLES ..................................................................................................................... 11
ABSTRACT ............................................................................................................................... 14
CHAPTER 1 INTRODUCTION ............................................................................................... 16
  STUDENTS’ AND TEACHERS’ PERCEPTIONS ................................................................. 18
  FOREIGN AND SECOND LANGUAGE TEACHING ......................................................... 23
  STUDENTS’ AND TEACHERS’ EVALUATIONS ............................................................... 27
  SUMMARY OF RESEARCH QUESTIONS ...................................................................... 28
  ORGANIZATION OF THE DISSERTATION ................................................................. 29
CHAPTER 2 REVIEW OF LITERATURE .............................................................................. 32
  EFFECTIVE SECOND AND FOREIGN LANGUAGE TEACHING .................................... 32
  BRIEF HISTORY OF FOREIGN LANGUAGE TEACHING METHODS ...................... 33
  CURRENT TOPICS IN FOREIGN LANGUAGE TEACHING ....................................... 38
  GRAMMAR TEACHING ................................................................................................. 39
  ERROR CORRECTION .................................................................................................... 50
  TARGET LANGUAGE USE ............................................................................................. 55
  CULTURE ......................................................................................................................... 63
  COMPUTER-ASSISTED LANGUAGE LEARNING ......................................................... 68
  COMMUNICATIVE APPROACHES TO LANGUAGE TEACHING ............................. 75
  ASSESSMENT .................................................................................................................... 87
  SECOND LANGUAGE LEARNING AND TEACHING STANDARDS ....................... 89
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDENTS’ AND TEACHERS’ PERCEPTIONS OF L2 TEACHING AND LEARNING</td>
<td>97</td>
</tr>
<tr>
<td>THE VALIDITY OF STUDENT RATINGS</td>
<td>111</td>
</tr>
<tr>
<td>STUDENT EVALUATIONS VS. TEACHER SELF-EVALUATIONS</td>
<td>119</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>123</td>
</tr>
<tr>
<td>CHAPTER 3 METHODOLOGY</td>
<td>124</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>124</td>
</tr>
<tr>
<td>PARTICIPANTS</td>
<td>126</td>
</tr>
<tr>
<td>GENERAL CONSIDERATIONS OF QUESTIONNAIRE DESIGN</td>
<td>130</td>
</tr>
<tr>
<td>PILOT TESTING</td>
<td>132</td>
</tr>
<tr>
<td>INSTRUMENTATION DEVELOPMENT AND CONTENT</td>
<td>134</td>
</tr>
<tr>
<td>PROCEDURES</td>
<td>157</td>
</tr>
<tr>
<td>DATA ANALYSIS</td>
<td>159</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>162</td>
</tr>
<tr>
<td>CHAPTER 4 DATA ANALYSIS AND RESULTS</td>
<td>163</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>163</td>
</tr>
<tr>
<td>STATISTICAL PROCEDURES</td>
<td>165</td>
</tr>
<tr>
<td>RESEARCH QUESTION 1</td>
<td>167</td>
</tr>
<tr>
<td>RESEARCH QUESTION 2</td>
<td>170</td>
</tr>
<tr>
<td>RESEARCH QUESTION 3</td>
<td>173</td>
</tr>
<tr>
<td>RESEARCH QUESTION 4</td>
<td>180</td>
</tr>
<tr>
<td>RESEARCH QUESTION 5</td>
<td>191</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS - Continued

RESEARCH QUESTION 6................................................................. 202
RESEARCH QUESTION 7................................................................. 211
RESEARCH QUESTION 8................................................................. 220
RESEARCH QUESTION 9................................................................. 223
SUMMARY .................................................................................. 226

CHAPTER 5 DISCUSSION AND CONCLUSIONS ......................... 227
INTRODUCTION ........................................................................... 227
IMPLICATIONS OF RESULTS BY CATEGORY AND QUESTIONNAIRE 228
TCE COMPARISONS .................................................................... 244
TEACHER COMPARISONS......................................................... 247
LANGUAGE AND LEVEL ............................................................. 252
DISCUSSION .............................................................................. 258
THEORETICAL AND PEDAGOGICAL IMPLICATIONS ................... 262
LIMITATIONS OF THIS STUDY .................................................. 265
RECOMMENDATIONS FOR FUTURE RESEARCH ....................... 269
CONCLUSION ............................................................................. 271

APPENDIX A: Effective Teacher Questionnaire ............................. 272
APPENDIX B: Student Information Questionnaire ....................... 273
APPENDIX C: Summary of Students’ Responses to Student Information Questionnaire
.................................................................................................................. 274
APPENDIX D: Teacher Information Questionnaire ........................ 276
APPENDIX E: Summary of Teachers’ Responses to Teacher Information Questionnaire ......................................................................................................................................... 277
APPENDIX F: Evaluation Questionnaire (Student Version) .................................................. 279
APPENDIX G: Evaluation Questionnaire (Teacher Version) ............................................. 280
APPENDIX H: TCE Questions .............................................................................................. 281
REFERENCES ......................................................................................................................... 282
LIST OF FIGURES

Figure 1. Bar Graph of Student Means on Effective Teacher Questionnaire. ............... 170

Figure 2. Bar Graph of Teacher Means on Effective Teacher Questionnaire ............. 173

Figure 3. Bar Graph of Student and Teacher Means on Effective Teacher Questionnaire.

........................................................................................................................................ 180
LIST OF TABLES

Table 1. Summary of Participating 1st and 2nd-Year Foreign Language Classes by Language and Level .......................................................... 129

Table 2. Summary of Factor Loadings on Items on Effective Teacher Questionnaire... 140

Table 3. Summary of Item Distribution across Logical Categories on Effective Teacher Questionnaire. ............................................................. 141

Table 4. Summary of Item Distribution across Categories on Evaluation Questionnaire. .................................................................................. 151

Table 5. Item and Abbreviated Reference Concordance. ........................................... 155

Table 6. Student Means on Effective Teacher Questionnaire ........................................ 169

Table 7. Teacher Means on Effective Teacher Questionnaire ................................. 172

Table 8. Correlations of Teachers’ and Students’ Responses by Item on Effective Teacher Questionnaire. .............................................................. 177

Table 9. Overall Comparison of Student and Teacher Means by Item on Effective Teacher Questionnaires including p Values from t-tests. .............. 178

Table 10. Summary of Averaged Differences between Individual Teachers’ Raw Scores and their Students’ Mean Responses and Direction of Difference by Item on Effective Teacher Questionnaire .......................................................... 184

Table 11. Averaged Teacher-Student Difference across 24 Items on Effective Teacher Questionnaire by Teacher .................................................. 188

Table 12. Pearson Product-moment Correlation for Individual Teachers and their Students on Effective Teacher Questionnaire ........................................ 190
LIST OF TABLES - Continued

Table 13. Overall Comparison and Correlations of Students’ and Teachers’ Responses on Frequency Column of Evaluation Questionnaire................................................................. 193

Table 14. Summary of Averaged Differences between Individual Teachers’ Raw Scores and their Students’ Mean Responses and Direction of Difference by Item on Frequency Column of Evaluation Questionnaire................................................................. 196

Table 15. Averaged Teacher-Student Difference across 21 Items on Frequency Column of Evaluation Questionnaire by Teacher........................................................................... 199

Table 16. Pearson Product-moment Correlation for Individual Teachers and their Students on Frequency Column of Evaluation Questionnaire.............................................. 201

Table 17. Overall Comparison and Correlations of Students’ and Teachers’ Responses on Evaluation Column on Evaluation Questionnaire................................................................. 203

Table 18. Summary of Averaged Differences between Individual Teachers’ Raw Scores and their Students’ Mean Responses and Direction of Difference by Item on Evaluation Column of Evaluation Questionnaire. ................................................................. 206

Table 19. Averaged Teacher-Student Difference across 21 Items on Evaluation Column of Evaluation Questionnaire by Teacher........................................................................... 209

Table 20. Pearson Product-moment Correlation for Individual Teachers and their Students on Evaluation Column of Evaluation Questionnaire................................................................. 210

Table 21. ANOVA of Language Type and Class Level by Item for Students’ Responses to Effective Teacher Questionnaire by Item. ................................................................. 214

Table 22. ANOVA of Language Type and Class Level by Item for Students’ Responses to Frequency Column on Evaluation Questionnaire................................................................. 216
LIST OF TABLES - Continued

Table 23. ANOVA of Language Type and Class Level by Item for Students’ Responses
to Evaluation Column on Evaluation Questionnaire............................................ 219

Table 24. Multiple Regression Analysis of TCE Question 1 and Students’ Responses on
All Questionnaires. .................................................................................................. 224

Table 25. Parameter Estimates for Multiple Regression Analysis. .......................... 224

Table 26. Multiple Regression Analysis of Mean of TCE Question 6 and Students’
Responses on All Questionnaires........................................................................... 225

Table 27. Parameter Estimates for Multiple Regression Analysis. ......................... 225

Table 28. Summary of Pearson Product-moment Correlations between Teachers and
Students by Teacher and Across Both Questionnaires......................................... 250
ABSTRACT

Although there seem to exist universal general principles of effective teaching, the foreign and second (L2) classroom presents learning objectives, tasks, and environments that are qualitatively distinct from those of other subjects. Relatively few studies have specifically compared and contrasted individual L2 teacher’s perceptions of effective teaching behaviors with those of their own students on similar instruments (Beaudrie, Brown, Thompson, 2004; Brosh, 1996, Kern, 1995a). As part of her dissertation, which addressed experienced teachers’ perceptions of effective L2 teaching, Reber (2001) suggests that “a similar study that would be valuable to the profession would be one comparing and matching teacher and student belief systems” (p. 173).

The current study explores L2 teachers’ and L2 students’ perceptions of effective L2 teaching by analyzing matches and mismatches between each group’s perception of what they feel an effective L2 teacher should know and be doing in the classroom. The concrete manifestation of these perceptions of effective teaching, in the classroom, is addressed via teaching evaluations. The principal objectives of this study were threefold: first, the identification and comparison of post-secondary L2 students’ and L2 teachers’ perceptions of effective teaching behaviors on a Likert-scale questionnaire; second, the comparison of students’ and teachers’ perceptions of how often specific teaching behaviors are performed; third, the comparison of students’ evaluations of teaching to their instructors’ self-evaluations on a similar questionnaire. A secondary objective of the study was to compare students’ responses on selected items from the university’s TCE form with their responses on the discipline-specific questionnaires used in this study.
Forty-nine teachers and their 1400 students from 83 intact beginning-level language classes (101-202) across nine distinct languages at the University of Arizona voluntarily participated in the study during Spring semester, 2005. During two visits, participating students and teachers filled out questionnaires regarding perceptions of 1) what effective FL teachers should be doing in the classroom, 2) how often certain target behaviors are performed, and 3) how effective teachers perform them. An additional component of the study involved the comparison of the students’ ratings on the language-teaching questionnaire with selected questions relative to teaching taken from the standard TCE form used university wide.

Statistical analyses, both descriptive and inferential, demonstrated that teachers and students, overall and by teacher, do have very different perceptions of what should be done in the FL classroom, what is currently being done, and how effectively it is being done. Questions that demonstrated statistically significant differences between teachers and students overall covered issues such as immediate error correction, task-based teaching, students’ use of FL early on, use of pair and small-group work, and grammar teaching. Teachers’ and students’ responses to the use of English in testing reading and listening skills, the need for the teacher to have native-like command of the target language, the simplification of the FL by the teacher to aid student understanding, and the necessity of situating grammar into real-world contexts were similar. Students and teachers seem to have differing opinions regarding grammar teaching and the usefulness of communicative language teaching strategies with students favoring a more traditional, grammar-based approach and teachers favoring a communicative FL classroom.
CHAPTER 1

INTRODUCTION

In the field of adult foreign (FL) and second language (L2) education the quandary of what exactly comprises effective teaching continues to perplex teachers, students, and administrators. This predicament is not isolated to the field of language teaching, as the ability to concretely identify specific teaching behaviors that are synonymous with learning eludes scholars in other fields as well. In their review of research on teacher performance criteria, Rosenshine & Furst (1971) admitted that “we know very little about the relationship between classroom behavior and student gains” (p. 37). Similarly, Travers (1981) proclaimed that so far a finite set of teacher competencies have not been identified to predict how much students will learn.

In spite of the difficulty in attributing specific learning gains to concrete teaching behaviors, certain general principles of effective pedagogy and complementary social variables appear to exercise some influence on all types of teaching and learning, regardless of discipline (Dunkin & Biddle, 1974; Murray, 1991; Rosenshine & Furst, 1971; Travers, 1981). In surveying the research on process-product studies, or studies relating teacher behaviors to student outcomes, Rosenshine & Furst (1971) point out eleven variables they believe exemplify effective teaching and, subsequently, result in effective learning:

1. Clarity (pp. 44-45)
2. Variability (pp. 45-46)
3. Enthusiasm (pp. 46-47)
4. Task-Oriented and/or Businesslike Behaviors (pp. 47-48)
5. Student Opportunity to Learn Criterion Material (pp. 48-49)
6. Use of Student Ideas and General Indirectness (pp. 49-50)
7. Criticism (pp. 50-51).
8. Use of Structuring Comments (pp. 51-52)
9. Types of Questions (pp. 52-53)
10. Probing (p. 53)
11. Level of Difficulty of Instruction (p. 54)

Travers (1981) took a holistic approach in arguing persuasively that variables within the teachers’ control make up only one aspect of the many issues to be considered when assessing student learning. He points to social variables such as the influence of the home environment and the extent of family members’ participation in and support of learning:

The extent to which a pupil learns in the school is a function of many different conditions, of which the teacher’s mode of operation is only one. Achievement is also a function of such conditions as the extent to which the home encourages learning, the extent to which the parents participate in the learning and development of the child, perhaps genetic endowment, and the kinds of material available for study (p. 18).

In examining college teaching, Murray (1991) reviews research on effective teaching behaviors, focusing on the impact of enthusiasm and clarity on student learning and motivation. Murray concedes that the majority of the research he analyzed was
conducted in lecture-style classes, but makes the bold assertion that “classroom teaching behaviors, at least in the enthusiasm and clarity domains, appear to be causal antecedents (rather than mere correlates) of various instructional outcome measures” (p. 161).

These authors’ claims that teacher behaviors are only one of many factors exerting influence on student learning make the improvement of teaching that much more crucial. If indeed there are many factors beyond teachers’ control that are affecting student achievement and learning, it becomes incumbent upon teachers to maximize the impact of those things they can control. Teachers cannot control students’ personal lives outside of the classroom, but they can control their own concrete pedagogical decisions and behaviors that directly impact students and their learning.

The general teaching principles cited above provide some guidance to teachers of all disciplines at an abstract level but lack the specificity needed to execute a concrete lesson plan that will, in turn, foster effective student learning. According to the authors cited above, it can be surmised that teachers need to be enthusiastic in their teaching, clear in their explanations, and liberal in their integration of student ideas—a rather intuitive proposition.

STUDENTS’ AND TEACHERS’ PERCEPTIONS

Teachers’ attitudes received much attention in the literature during the early 1950’s and early 1970’s and, more recently, they have resurfaced as key to understanding what motivates teachers’ actions (Richardson, 1996). Among other terms, Richardson (1996) groups attitudes, beliefs, and perceptions as a set of mental constructs that “name, define, and describe the structure and content of mental states thought to drive a person’s
actions” (p. 102). She also alludes to the definition offered by anthropologists, social psychologists, and philosophers that considers beliefs as psychological propositions, premises, and other understandings about the world that are felt to be true (Richardson, 1996).

Rather than alternate between various terms and to facilitate discussion on the topic, the term ‘belief’ will be used in reviewing literature in the area of teachers’ and students’ beliefs regarding ideal practices. ‘Perception’ will be used in discussing the particulars of this study, which includes beliefs as defined above as well as students’ subjective evaluations of their teachers’ behaviors. Hence, for the purposes of this study ‘perception’ will be operationalized to mean participants’ psychologically held, subjective beliefs on ideal teaching practices, their observations of the frequency of specific teaching practices, and their assessments of those practices. Simply put, ‘perceptions’ within the framework of this study’s research design refers to students’ and teachers’ personal, subjective beliefs on three different levels: 1) which teaching practices students and teachers generally believe to be effective in foreign language teaching, 2) how often they perceive certain behaviors to be occurring in their classrooms, and 3) how effectively they believe their teachers perform certain behaviors. These three levels closely parallel a three-part definition explicated by Pajares (1992) where beliefs may be prescriptive, descriptive, and evaluative respectively. The conflation of terms should not be construed as an attempt to minimize the importance of the subtle semantic differences between beliefs, attitudes, conceptions, perceptions, and perspectives; however, given the
purpose and scope of the current study, ‘perceptions’ adequately defines the
individualistic, subjective nature of this study’s target construct.

Arguing in favor of a social constructivist approach to language learning—both
naturalistic and instructed—Williams and Burden (1997) also bring to the forefront the
close relationship between beliefs and actions on the part of both teachers and students.
They propose that teachers’ actions are highly influenced by their beliefs and assert that
“one of the many facets that teachers bring to the teaching-learning process is a view of
what education is all about, and this belief, whether implicit or explicit, will influence
their actions in the classroom” (pp. 48-49). The authors propose that an essential stage in
the social constructivist approach is for teachers to identify what their beliefs are.

Although students’ and teachers’ personal beliefs regarding teaching may not be
corroborated by experimental or scientific evidence or be perfectly predictable, there
appears to be a strong relationship between previous experience and the development of
notions relative to teaching and learning. In examining teachers in training and their
beliefs regarding teaching, Richardson (1996) identifies the following three types of
experience as influential: personal experience, experience with schooling and instruction,
and experience with formal knowledge. Personal experience has reference to one’s
worldview, values, intellectual persuasions, moral tendencies, and familial and cultural
background. Experience with schooling and instruction refers specifically to teachers’
previous experiences when they were students in traditional classrooms. These first two
types of experience result in very deeply rooted ideas concerning teaching and learning
that may be difficult to alter. Experience with formal knowledge includes experience with
understandings that have been agreed upon as credible and valid by a community of scholars. Students engage with formal knowledge through various means including information presented through the mass media in the form of news and documentaries or articles in magazines or journals.

In his review of the belief literature, Pajares (1992) confirms Richardson’s claim that, (1) beliefs about teaching are formulated through many years of formal schooling, and, (2) that they may be difficult to alter. He cites Lortie (1975) who used the term “apprenticeship of observation” in referring to the impact of previous experience in the molding of educational beliefs. Pajares argues that teachers’ tendency to enter the profession with preconceived notions of what teaching should entail contrasts with other professions such as medicine and law. Preservice teachers can be considered insiders and as such do not feel the need to redefine their situation. These aspiring teachers’ reality and beliefs remain largely unaltered by courses in higher education. He claims that beginning medical and law students enter their respective fields and places of employment, e.g., the operating room or the courtroom, with little to no exposure or experience—unlike teachers. Specifically, Pajares alleges that for medical and law students the operating room and the courtroom are “new to students, what goes on in them is alien, and understandings must be constructed nearly from scratch” (p. 323). Beliefs, according to Pajares (1992), can become so entrenched that often they are maintained even after having been shown to be empirically incorrect and unrepresentative of reality.
Even though teachers’ belief structures may be extremely resistant to change, Richardson (1996) argues that change can indeed occur. In her summary of the research on studies in belief change, the author notes that in-service teachers tend to be more open to change than pre-service teachers. Another catalyst for belief change resides in a constructivist approach where teachers are encouraged to engage in reflective examination of their practices and beliefs (Richardson, 1996).

Equally important in improving teaching, student learning, and student achievement is a firm understanding of students’ beliefs and perceptions. Williams and Burden (1997) claim that “learners’ perceptions and interpretations . . . have been found to have the greatest influence on achievement” (p. 98). Not surprisingly, they maintain that in some cases students’ perceptions of teacher behaviors do not correspond with teachers’ intentions (Williams & Burden, 1997).

Students and teachers may have very similar or disparate notions of effective teaching and the intersection of the two has ramifications for students’ language learning and the effectiveness of teachers’ instruction. E.K. Horwitz (1990), one of the leading scholars on the affective domain in the foreign language classroom—specifically issues of anxiety (Horwitz, 1991)—argues that mismatches between students’ and teachers’ expectations of teaching “can lead to a lack of student confidence in and satisfaction with the language class” (p. 25). Likewise, Kern (1995a) summarizes the purpose of research into students’ and teachers’ beliefs about language teaching by stating that the main goal is “to predict expectational conflicts that may contribute to student frustration, anxiety, lack of motivation, and in some cases, ending of foreign language study” (p. 71). It is
noteworthy that the aforementioned claims regarding student frustration and lack of motivation were made without reference to any sort of concrete, quantifiable learning gains or progress in achievement, i.e., grades. The authors cited above express concerns relative to more far-reaching consequences, namely, disappointment with learning and disillusionment with formal education, their teachers and the institutions they represent.

The simple fact that for some students a mismatch in expectations for classroom teaching may result in disillusionment, regardless of achievement or grades, warrants further inquiry into the nature of and correspondence between students’ and teachers’ expectations of concrete teaching practices. This dilemma is further complicated by the complex and unique nature of language learning and teaching that will be outlined below.

FOREIGN AND SECOND LANGUAGE TEACHING

In addition to the generally accepted principles of effective teaching noted above, the L2 classroom presents learning objectives, tasks, and environments that are qualitatively distinct from those of other subjects. Reber (2001) claims that “several aspects of FL teaching are distinctly different from teaching in other disciplines” (p. 33). Unlike content classes in a first language (L1) where the transmission of conceptual knowledge and concrete facts takes place by way of a mutually intelligible language, L2 learning not only includes the transmission of concepts and facts but does so, in many cases, via the very subject under examination—the second language. Similarly, Brosh (1996) notes that in the communicative L2 classroom the “means of instruction is also the subject of instruction” (p. 125). The unique nature of human language learning as compared to other types of learning has led to the proposal of a language acquisition
device (LAD) that is innate to the human brain by Noam Chomsky (1965). According to Chomsky, the ability for children to acquire such a complex skill as their first language can only be explained by a unique human ability to acquire languages, i.e., the language acquisition device. While the acquisition of other skills and content knowledge may be cognitively challenging and attributable to non-species specific learning mechanisms, the acquisition of language presents a qualitatively distinct type of learning that the existence of the LAD helps explain.

In the modern L2 classroom the very means of communication becomes the subject matter, and the need to manipulate and comprehend foreign sounds, unfamiliar words, abstract grammatical structures, and cultural nuances makes the task formidable. Hence, in order to accurately define effective L2 pedagogy one must weigh in on fundamental and controversial issues such as target-language use, grammar instruction, error correction, language-culture connections, and communicative language teaching—all of which reveal one’s basic understanding of not only how to learn another language but what language is.

Language learning presents cognitively challenging tasks that also have social, psychological, and interpersonal consequences as well. The unique nature of learning a code of communication of a different community of people has ramifications for self-identity, cultural identity, and self-image. This idea is elucidated by Williams and Burden (1991) in the following manner:

Language, after all, belongs to a person’s whole social being; it is a part of one’s identity, and is used to convey this identity to other people. The learning of a
foreign language involves far more than simply learning skills, or a system of rules, or a grammar; it involves an alteration of self-image, the adoption of new social and cultural behaviors and ways of being, and therefore has a significant impact on the social nature of the learner (p. 115).

According to the teacher belief literature reviewed above, the way in which language teachers choose to instruct their students is most likely a reflection of their belief system and reveals their understanding of how language is fundamentally represented, accessed, and, ultimately, acquired. For example, a teacher who speaks nothing but the target language with little to no overt grammar instruction appears to believe that a second language is learned, acquired, and accessed similar to a first language. The teacher who focuses on explicit, deductive grammar instruction of the L2 via the L1 seems to assume that through contrastive analysis with the first language, accuracy and understanding can be attained. Not only do many language instructors have certain notions of what constitutes effective language teaching and learning, so too do many of their students (Brosh, 1996; Horwitz, 1988; Kern, 1995a; Levine, 2003; Reber, 2001; Schulz, 1996; Schulz, 2001; Wennerstrom & Heiser, 1992).

While the literature is replete with studies delving into the perceptions of learners and teachers concerning different aspects of language teaching and learning (Brosh, 1996; Horwitz, 1988; Levine, 2003; Reber, 2001; Schulz, 1996; Schulz, 2001; Wennerstrom & Heiser, 1992), relatively few have specifically compared and contrasted individual teachers’ perceptions of effective teaching practices with those of their very own students (Kern, 1995a). Many studies have explored the perceptions of one group,
either teachers or students, but have excluded the other’s (Harrington & Hertel, 2000; Horwitz, 1985; Horwitz, 1988). Those studies that have included both groups’ perceptions have not been conducted or reported in such a manner as to allow for direct comparisons between what Teacher ‘X’ believes and the students of Teacher ‘X.’

Another issue central to teacher and student belief research is the use of questionnaires and their rate of response. Survey research is notorious for resulting in low response rates, especially with mailed questionnaires (Brown, 2001). Dornyei notes that a response rate of 50% on mailed questionnaires is acceptable while 80% is outstanding (2003). When a participant is allowed to complete the questionnaire on their own, removed from the presence of the researcher, response rates will inevitably decrease. Even with the added convenience of on-line questionnaires, response rates may still suffer. Levine (2003) sent out emails to foreign language instructors and program directors across the United States and Canada requesting their participation, and their students’, on an anonymous, on-line questionnaire. This procedure resulted in 163 responses by instructors and 600 by students, an average of 3.7 students per instructor. Clearly, an on-line approach to questionnaire research is advantageous for reasons of convenience in data collection. However, Levine’s design did not allow for direct comparisons between individual teacher’s responses and their specific students as the two were not connected. Even if students were connected to specific teachers in this study, the statistical analysis would not have been robust with an average of only 3.7 students per instructor. Face-to-face administration within the context of the classroom and during class time, which Dornyei (2003) calls “group administration” (p. 82), would seem to
result in higher response rates for very practical and logistical reasons. Dornyei (2003) claims that this approach may even result in a return of 100% of questionnaires.

STUDENTS’ AND TEACHERS’ EVALUATIONS

To this researcher’s knowledge very little research exists within L2 language teaching that has directly compared L2 teachers’ self-assessments with the evaluations of their own students using a similar evaluation instrument (Beaudrie, Brown, & Thompson, 2004). Feldman (1989) reviewed research conducted across several fields where multiple sources of teaching evaluation were included. Unfortunately, his analysis did not separate by field nor was much information given on the similarity of instruments used by each evaluation group. This meta-analysis of several studies and their respective research designs will be considered in more depth in Chapter 2. As part of her dissertation which addressed experienced teachers’ perceptions of effective L2 teaching, Reber (2001) makes recommendations for future research in the area of teachers’ and students’ perceptions of effective teaching. In one such recommendation, she states, “a similar study that would be valuable to the profession would be one comparing and matching teacher and student belief systems” (p. 173).

The current study attends to Reber’s (2001) suggestion by exploring foreign language teachers’ and students’ perceptions of effective FL teaching by analyzing matches and mismatches between each group’s perception of effective teaching behaviors. The concrete manifestation of these perceptions of effective teaching in the FL classroom will also be addressed vis-a-vis teaching evaluations. The relationship between L2 students’ evaluations of their instructors’ teaching and those same instructors’ self-
evaluations will be explored. It is with these objectives that the following research questions have been formulated.

SUMMARY OF RESEARCH QUESTIONS

Research Question 1

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do students consider to be reflective of effective foreign language teachers?

Research Question 2

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do these students’ teachers consider to be reflective of effective foreign language teachers?

Research Question 3

Based on a similar 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), how do students’ and teachers’ perceptions of effective foreign language teachers coincide or differ?

Research Question 4

How do individual teachers’ perceptions of effective foreign language teachers coincide with or differ from those of their students’?

Research Question 5

How do students’ and teachers’ perceptions of the frequency of occurrence of certain teaching behaviors in their classrooms coincide or differ, both overall and in a
teacher-by-teacher comparison on a 21-item, Likert-scale questionnaire (Evaluation Questionnaire)?

Research Question 6

How do teachers’ self-evaluations coincide with or differ from their students’ evaluations of them overall and individually?

Research Question 7

What impact do language type and class level exert on students’ responses to both the 24-item (Effective Teacher) and 21-item (Evaluation) questionnaires?

Research Question 8

How do students’ evaluations of their foreign language teachers on the 21-item language-specific instrument (Evaluation Questionnaire) coincide with or differ from their evaluation of their teachers on selected items from the general, university-wide evaluation form, i.e., the Teacher-Course Evaluation (TCE)?

Research Question 9

What relationship exists between students’ response patterns on both the Effective Teacher and Evaluation Questionnaires and students’ responses to selected items from the TCE instrument?

ORGANIZATION OF THE DISSERTATION

Chapter 1 has briefly introduced the reader to topics considered central to the current research project by presenting background information on the general issue of teaching effectiveness, the relationship between teachers’ beliefs and actions, the unique nature of foreign and second language learning, and, finally, the need for more detailed,
comparative analyses of FL students’ and teachers’ perceptions of teacher behaviors.

Chapter 2 will survey the theoretical and empirical literature in several areas within foreign language pedagogy deemed relevant to the current research project and the design of the study’s questionnaires. The first area will treat the notion of what currently constitutes effective foreign language teaching from different perspectives. A brief historical outline of language teaching methodologies will be included, followed by a discussion of the recent movement to generate professional teaching standards for practicing teachers and learning standards for students. Applicable empirical research and theoretical propositions that have recently influenced the field of SL and FL teaching will also be traced and related to this study’s instrumentation. Second, students’ and teachers’ perceptions regarding foreign language teaching and learning will be presented with a primary focus on those issues included in the current instruments. Third, the validity of student ratings of teaching will receive attention followed by a comparison of students’ evaluations to teachers’ self-evaluations.

Chapter 3 will begin by restating the research questions before delineating the specifics of the research context, the development of the instrumentation and rationale behind the inclusion of each item, and the data collection and analysis procedures. Chapter 4, then, will present the results of the study as they pertain to the stated research questions with some preliminary analysis and discussion. The final chapter, Chapter 5, will summarize the results of the study, provide analysis, and discuss implications of the results for the field at large. Moreover, Chapter 5 will address the limitations of the study and research design while also recommending future research relative to students’ and
teachers’ belief systems, their match or mismatch, and the impact those beliefs may have on pedagogical practices in the foreign language classroom.
CHAPTER 2
REVIEW OF LITERATURE

EFFECTIVE SECOND AND FOREIGN LANGUAGE TEACHING

The ever-changing nature of L2 acquisition and pedagogical theories and pedagogy makes it difficult to narrow down one methodology or approach that would embody effective L2 teaching in all contexts. Second language learning theories frequently rise and fall from favor along with their accompanying methodologies which complicates the issue of what effective language teaching may, or may not, entail. Mitchell and Vidal (2001) argue against the utility of the oft-used pendulum metaphor in describing the evolution of L2 theory and practice. They liken the dynamic nature of language learning theories and methods to the ebbs and flows of a methodological river with small creeks and streams flowing into the main river of L2 pedagogy contributing to its overall direction. Therefore, the search for a concise description of effective teaching behaviors in the L2 classroom must take into account theoretical and practical considerations backed by empirical and experiential evidence.

In the first chapter of their book, *Approaches and Methods in Language Teaching*, Richards and Rodgers (2001) offer a brief history of language teaching and trace the rise and fall of diverse approaches and methodologies. An abbreviated summary will be included here to provide needed background on the evolution of teaching methods and how methodological shifts reflect concurrent thinking on proficiency needs as well as theories of language learning. These shifting trends in language pedagogy are often influenced by theories and findings from other fields such as educational and social
psychology. In many cases issues that are being grappled with currently by the field, i.e.,
grammar instruction, have been the focus of attention by earlier pedagogical approaches,
 i.e., audio-lingual method, albeit in different ways. Given the present study’s focus on
teaching behaviors, the following review of language teaching methodologies and
approaches will pay particular attention to the role of teachers and the preferred teaching
practices as prescribed by each methodology.

BRIEF HISTORY OF FOREIGN LANGUAGE TEACHING METHODS

Richards & Rodgers (2001) begin their historical overview with the teaching of
Latin and Greek and its tremendous influence on language pedagogy that began 500
years ago. From the fifteenth through the nineteenth century language students were
schooled in the particulars of Latin through rote memorization of grammar and sample
sentences. Teachers need not be conversational or truly communicative in the language
rather able to explain the grammar in detail and model it through stilted, decontextualized
sentences. The expert teacher was primarily a grammarian with extensive knowledge of
Latin grammar. Even with the rise of modern languages such as French and English and
the decline of Latin as a common means of spoken and written communication, Latin
continued as a necessary academic subject. Many viewed the study of Latin as crucial for
intellectual development.

The teaching of Latin via rote memorization and direct translation of
decontextualized sample sentences continued from the fifteenth century on through the
nineteenth century and was embodied by the ‘Grammar-Translation’ method. Even after
modern languages entered the school curriculum, the Grammar-Translation method
prioritized the reading of a language’s literature and deemed it a valuable mental
exercise. Reading and writing were the skills of choice, accuracy was paramount,
grammar was approached deductively, and students’ native language was the medium of
instruction. It was essential for teachers to have a thorough mastery of a language’s
grammar and the ability to effectively police students’ production, specifically grammar.
This method continued strong on into the twentieth century (Richards & Rodgers, 2001).

A reaction against the Grammar-Translation movement surfaced which held up
the successful acquisition of language by children as a model for second language
learning. The reform movement that arose emphasized the spoken word with its sound
system and contextualized nature. Out of this reformation grew the Natural Method,
which attempted to mirror first language acquisition without recourse to the native
language during instruction. One of the most well known of the natural methods was the
Direct Method which was adopted by Berlitz in his commercial schools and led to the
term ‘Berlitz’ method being used synonymously with the Direct Method. The Direct
Method focused on oral and aural skills with the entire class being taught in the target
language with correct pronunciation and grammar considered paramount. Grammar was
taught inductively and vocabulary was presented through gestures, objects, and pictures.
(Richards & Rodgers, 2001).

Although the Direct Method seemed to improve upon the many shortcomings of
the Grammar-Translation method it too had several drawbacks. Teachers had to
demonstrate native, or native-like, fluency in the target language and they needed to be
extremely skillful in their ability to communicate ideas and teach vocabulary using
gesture and body language while maintaining a high level of energy and motivation among students. Also, the strict use of the target language to the exclusion of the native language resulted in teachers having to perform time consuming verbal gymnastics. The possibility of implementing such a method in the public schools was impractical and far too ambitious given the great demands put on the teacher with large class sizes. The Coleman Report of 1927 recommended that a more achievable goal for foreign language programs was the development of reading knowledge through simple texts (Richards & Rodgers, 2001).

The Direct Method, however, opened the door to the methods era with its accompanying assumptions, namely, 1) that approaches and methods refer to a consistent set of teaching strategies employed by the teacher that are theoretically based, 2) that certain methods or approaches will lead to more effective learning, and 3) that quality teaching will improve learning when the best methods are used. From the time directly following WWII (1950’s) through the 1980’s, several different methods surfaced as a result of different goals and shifts in linguistic and educational theories—all based on the assumptions listed above (Richards & Rodgers, 2001).

The Audiolingual Method surfaced as a way to afford students oral communication skills that the Grammar-Translation method could not. Furthermore, it was a logical by-product of the advances in behaviorist psychology that advocated stimulus-response approaches to learning. As a result of the Cold War and the launching of Sputnik in 1957 by the Russians, the United States government sought to match the Russians advances in technology and felt that increased efforts to improve science, math,
and language learning in schools would help bridge the gap. Hence, the government
invested heavily into FL language teaching in order to improve American students’
foreign language abilities throughout the country. These language labs were characterized
by a teacher-controlled console with individual student stations equipped with
audiocassette players, recorders, and headphones. From their console teachers could play
for all students at the same time certain sound clips for students to repeat as practice or
subsequently speak into their recorder. The teacher, therefore, was required to produce
native-like language both phonologically and grammatically while being able to detect
specific errors as well as error patterns among all students. In spite of this method’s
popularity for years, its effectiveness began to be called into question. Students
memorized set phrases and were required to articulate them flawlessly, but they had no
communicative ability and resembled programmed, language robots.

The 1960s, 1970s, and 1980s saw the birth of other methods and approaches,
some of which had larger followings than others. The Silent Way, the Natural Approach,
Total Physical Response, Suggestopedia, Counseling-Learning, and Communicative
Language Teaching all offered their ideas and prescriptions on how best to teach
language and what was the role of the teacher. Some of these approaches assigned a more
passive role to teachers, who served as counselors and interpreters to be used at the
students’ discretion as set out by the Silent Way, Suggestopedia, and Counseling-
Learning approaches. The Natural Approach, Total Physical Response, and
Communicative Language Teaching required a more active role of the teacher as
facilitator and instigator of language learning and interaction. While Task-Based
Language Teaching, Content-Based Instruction, and Proficiency-Based approaches focused on the outcomes of learning and not necessarily on the methods, they did require the teacher to be actively involved and carefully selective in their choice of tasks, content, and activities to which their students should be exposed. Multiple intelligences, Cooperative Learning, and Whole Language are trends from outside the field of second language learning that have been related to language pedagogy and require a tremendous amount of awareness and knowledge on the part of the teacher regarding the make-up of their students’ abilities and learning preferences. (Richards & Rodgers, 2001).

Richards and Rogers (2001) suggest that currently the language teaching profession is in a ‘post-methods’ era, arguing that no longer should teachers feel obligated to follow the overly prescriptive mandates of one certain method or approach. They encourage teachers to identify principles of effective language teaching that would guide their classroom decisions, not dictate them. Some of these crucial principles that they espouse relate directly to empirical work done in the field and include the following principles:

- Engage all learners in the lesson.
- Make learners, and not the teacher, the focus of the lesson.
- Provide maximum opportunities for student participation.
- Develop learner responsibility.
- Be tolerant of mistakes.
- Develop learners’ confidence.
- Teach learning strategies.
- Respond to learners’ difficulties and build on them.
- Use a maximum amount of student-to-student activities
- Promote cooperation among learners.
- Practice both accuracy and fluency.
- Address learners’ needs and interests (Richards & Rodgers, 2001, p. 251)

These principles provide some guidance, but teachers still must make concrete decisions on which activities they will implement in teaching their students within the context of their own classrooms. Armed with this information, then, teachers must choose which behaviors, on their part, will be most effective in stimulating second language acquisition.

CURRENT TOPICS IN FOREIGN LANGUAGE TEACHING

Unfortunately, given the multifaceted nature of second language acquisition and its equally complex relationship to L2 pedagogy, there is still little consensus on what exactly makes up effective L2 teaching and little empirical evidence directly linking certain teaching behaviors with effective SLA. Different research directions in SLA that informed the design of the current study’s instrumentation have been followed by various researchers and have appeared productive at times for not only theory building in SLA but also praxis. Several issues in foreign and second language learning that have relevance for the current study are included in this review. Much of this research is based on empirical findings by well-known researchers in the field of second language acquisition and second language pedagogy. Through their research findings SLA scholars have contributed to our understanding of the extremely complex and multi-faceted
enterprise that L2 acquisition represents. Nevertheless, much remains to be learned about how to integrate these findings into concrete practice.

The purpose of this review is to establish the importance of each area in current thinking on L2 pedagogy in order to justify the inclusion of related items on the study’s questionnaires. Since the primary objective of the current study is an exploratory comparison of teachers’ and students’ opinions of L2 pedagogical practices in the classroom, the intent is to establish their relevance—not to prove or disprove them. Likewise, the aim is not to provide an exhaustive list of current approaches to L2 pedagogy, but rather a sampling of practical pedagogical considerations reflected in the current study’s instrumentation. Hence, both professional consensus and the review of literature below validate the inclusion of the items that appear on the questionnaires. The following topics in L2 pedagogy appear on the instrumentation and thus will receive attention in turn; they are grammar teaching, error correction, target language use, culture teaching, computer-based language learning, various approaches to communicative language teaching, and L2 assessment.

GRAMMAR TEACHING

Arguably one of the most controversial and extensively debated issues in the area of adult SLA is the question of naturalistic versus instructed language learning. This debate revolves around the effectiveness of providing explicit instruction to adults learning another language or leaving them to their own resources in acquiring a second language. Many related questions in SLA that have received much attention in the research literature stem from this central preoccupation with the role of instruction. Some
of the most significant are the degree of access to Universal Grammar, the existence of a critical, or sensitive, period for language learning, and the validity of the learning-acquisition dichotomy put forth by Krashen (1982). Those who espouse a strong ‘non-interventionist’ approach to SLA question the need for instruction in learning a language and, consequently, claim that adult learners have full access to Universal Grammar. Scholars in this camp argue that positive evidence given via authentic input is sufficient for SLA to take place without resorting to negative or explicitly presented evidence (Felix, 1981; Schwartz, 1993; Schwartz & Sprouse, 1996). As a result, those who make a case against L2 instruction are less likely to agree with the notion of a critical period for language learning.

On the other side of the continuum reside the majority who find L2 instruction facilitative in achieving adult SLA and who believe that there is a critical period for language learning. Early studies in this research area were triggered by Krashen’s (1982) well-known learning/acquisition hypothesis and attempted to verify or refute Krashen’s claims that conscious learning of language and its forms did not contribute to SLA. Among others, Long (1988) was one of the pioneers in the field who analyzed the role of instruction in SLA and identified four domains affected by instruction: 1) SLA processes, i.e., transfer, generalization, elaboration, stabilization, 2) SLA route, i.e., developmental sequences of acquisition for certain forms, 3) SLA rate, i.e., the amount of time needed to acquire a certain form and pass from one stage to another, and, 4) level of ultimate SL attainment, i.e., degree of approximation to native-like mastery of the target language. Mounting evidence in each of these domains led Doughty (2003) in her review of the
literature on instructed SLA to conclude that “by the 1990’s, the evidence in the four domains of SLA, although scant, formed the basis of an assumption that L2 instruction is effective” (p. 263).

Once the general question of whether instructed SLA had a positive impact was answered it followed to discover which manner of instruction was best. Doughty (2003) specifies two general approaches as key in studying the effect of instruction—explicit and implicit. Explicit instruction refers to teaching strategies where rules are explained to learners (deductive) or where learners are overtly told to discover rules by attending to form (inductive). Implicit instruction makes no overt reference to rules, nor explicitly asks students to attend to forms. Long & Robinson (1998) refer to three instructional options in language teaching: focus on form, focus on formS, and focus on meaning. A focus on form curriculum treats grammar as it arises incidentally in a classroom based around meaningful, communicative language. This approach is responsive in nature as it reacts to the grammatical structures students concretely struggle with and in the order they arise. A focus on formS pedagogy constructs its syllabus around certain discrete, linguistic features that are added in a linear fashion. This approach is proactive in grammar instruction with a prepared list of structures and when they will appear in the curriculum. Those who espouse a focus on meaning approach generally ascribe to the notion that L1 and L2 acquisition occur without explicit focus on language per se, rather on communication. This program relies on positive evidence and claims that incidental and implicit learning of grammar suffice. Within this approach, any focus on language must be natural and meaningful.
Others have offered taxonomies for categorizing pedagogical interventions for teaching linguistic forms including Norris and Ortega (2000) and Doughty and Williams (1998). In their review of studies into instructed SLA, Norris and Ortega classified each according to whether they were implicit or explicit and whether they focused on meaning, forms, or form. Doughty and Williams distinguished various degrees of focus on form placing different instructional types, i.e., input enhancement, negotiation, recast, output enhancement, etc., along an attention-to-form continuum ranging from unobtrusive to obtrusive. They considered the Garden Path method as most obtrusive since teachers induce students to make overgeneralization errors to then provide explicit error correction, e.g., she walks > she walked, she types > she typed, she goes > she goed, and input flood and task-essential language as most unobtrusive as these last two techniques do not make any effort to correct students or even point out grammatical structures.

After Long’s initial rekindling of interest in the area of instructed SLA in the early 1980’s and possibly as a reaction to the overwhelming integration of communicative approaches to L2 pedagogy which struggled to produce accurate speakers, many scholars have explored the effectiveness of explicit grammar teaching in the L2 classroom. Doughty and Williams (1998) edited a multi-chapter volume dedicated specifically to the study of form-focused instruction entitled *Focus on Form in Classroom Second Language Acquisition*. Three representative studies of the focus on form tradition which were conducted by Rosa and O’Neill (1999), Lightbown and Spada (1990), Tomasello and Herron (1988, 1989, 1991) are explained below.
Rosa and O’Neill (1999) investigated the relationship between the explicitness of instruction, the level of student awareness, and the degree of intake in native English speakers enrolled in fourth-semester Spanish. A total of 67 students participated in the study with seventy-one students being eliminated for not meeting the participant criteria, which included pretest scores, native language, incomplete or defective protocols. The researchers chose the Spanish contrary-to-fact conditional sentences in the past as the target structure. These five groups completed a multiple-choice jigsaw task as the method of presentation. Each student received a packet containing 10 pages. Each page had pasted on it (a) a piece of the puzzle depicting an event and (b) another piece of the puzzle with the main clause of the contrary-to-fact sentence type. Three other loose puzzle pieces which contained possible subordinate clauses were included with each page. Of these subordinate-clause puzzle pieces, only one fit perfectly with the main-clause puzzle piece.

The participants were randomly divided into five groups which represented different teaching and task direction combinations. In completing the task, each group received different instructions. The rule search [+RS] groups received explicit instructions to try and think of rules that explained the correct combination of clauses. The [-RS] groups were instructed to try and memorize the information in the picture since they would be asked about the people they saw later on in the study. All groups received an answer sheet; three groups were asked to write down the correct clause after completing each puzzle, while the other two groups responded to ten questions regarding content and word order. Although all groups received instruction on contrary-to-fact
sentences in a present time frame prior to the experiment, only two groups [+FI] received formal instruction on conditional sentences in a past time frame by way of a packet of materials containing grammatical explanations. The following four groups resulted from the combination of the aforementioned tasks: [+FI, +RS], [+FI, -RS], [-FI, +RS], [-FI, -RS], [T-only]. Finally, a fifth group, considered the control group, did not receive any explicit instructions concerning how they should approach the puzzle task.

The results of Rosa and O’Neill’s (1999) study showed that the two groups that received formal instruction, [+FI, +RS] and [+FI, -RS], significantly outperformed the group that did not receive explicit instruction, [-FI, -RS], from the pretest to the posttest. The students who completed the puzzle task without any explicit instruction, [-FI, -RS], evidenced an inferior understanding of the target structure. Finally, formal instruction appeared to overpower the distracting effect of the memorizing task as the group that received formal instruction [+FI,-RS] outperformed the group that did not [-FI, -RS], even though both participated in the memorization task.

Regarding the issue of awareness, the researchers state, “awareness at a higher level, namely at the level of rule formation, was a strong contributor to pretest-to-posttest improvement than lower levels of awareness, be it awareness expressed as explicit mention of verb forms ([+N]) or no explicit report of being aware ([NVR])” (Rosa & O’Neill, 1999, p. 540).

Lightbown and Spada (1990) found similar results on explicit instruction in attempting to answer the question “To what extent is form-focused instruction beneficial to classroom learners of a second language?” (p. 430). Observational data were collected
from four ESL classes comprised of 10 to 12-year-old native French speakers enrolled in a five-month intensive course. The researchers examined listening and reading comprehension skills as well as oral production skills. In addition, they analyzed the amount of contact the students had with English speakers outside the classroom and their attitudes toward the target language. Finally, scholars closely examined the type of instruction students received.

Although all four classes generally reflected a communicative approach with classroom activities focused on meaning rather than form, teachers differed greatly in the amount of time spent on form-focused activities. While 29 percent of total observed time in class 1 was devoted to form, the teacher of class 3 spent only 11 percent of her instruction time on form. Each teacher seemed to have a different grammatical focus, and the researchers hypothesized that students’ performance on those specific structures would be affected. Analysis of students’ grammatical performance showed that students in class 1 produced more accurate forms, especially in reference to the -ing ending, there is vs. you have construction, and possessive determiners. Students in class 4, who received the least form-focused instruction, had the lowest level of accuracy on all grammatical features examined. Nevertheless, students in class 4 did not lag behind in their overall communicative performance, including their comprehension skills (Lightbown & Spada, 1990). Lightbown and Spada’s findings are quite logical. Teachers who focus on certain forms in their interaction with students will most likely cause them to be more sensitive to those structures, and thus students will experience consciousness raising.
White, Spada, and Ranta (1991) conducted a study about the effectiveness of input enhancement activities on English question formation with beginning level francophone ESL learners. The authors based their study on the assumption that L2 learners will inevitably fail to perceive certain structures even after receiving exposure to naturalistic discourse. Similar to negative feedback, input enhancement attempts to draw the learners’ attention to specific properties of the input that they might otherwise overlook. In phase I of the study, two classes (n=53) received form-focused instruction specifically on question formation, while three other classes (n=76) did not receive instruction on questions, but rather on adverb placement. The first day after termination of the instruction period, students performed a question-forming task that had them reorder words to reflect correct English question word order. The resulting data from the tests showed that those who received explicit instruction and continuous corrective feedback in question formation from their instructors outperformed the subjects from the control group. Although not statistically significant, the group receiving instruction formulated correct questions 55 percent more often than their counterparts who did not receive instruction.

Phase II of White, Spada, and Ranta’s study (1991) included an oral component that required students to formulate questions in order to match a picture given them with a hidden picture held by the test administrator. Once again, the two groups differed in that one received overt instruction on question formation and corrective feedback, while the other group was instructed on adverb placement. In a comparison of the pre-, post-, and follow-up tests, the results demonstrate that the learners did benefit from the
instruction, as 35 students showed a 10 percent increased accuracy from pretest to follow-up; 15 students were within 10 percent of their pretest scores; and only 3 students’ accuracy fell by more than 10 percent.

The increased syntactic ability of the students to form questions points to the validity of explicitly teaching language structures accompanied by monitoring. The researchers also conclude that their study showed an influence in the “‘students’ oral performance in relatively spontaneous or ‘unmonitored tasks.’” They state, “We take these changes in oral performance as evidence that input enhancement can bring about genuine changes in learners’ interlanguage systems” (White et al., 1991, p. 429).

Although White, Spada, and Ranta’s (1991) study demonstrates the benefit of coupling explicit grammar instruction with negative feedback, it shows no benefit of using just one or the other. This study would be much improved had the researchers included two other experimental groups: a group that would receive only negative feedback and a second group that would receive only overt grammar instruction. Gains made by the experimental students in the above-mentioned study might have been much less had they not received both treatments.

From the theoretical base of consciousness-raising and form-focused instruction, Tomasello and Herron (1988) conducted a study that compared traditional deductive instruction, Error Avoidance, with the Garden Path Method. This latter approach induces students to overgeneralize grammatical principles in order to help them heighten their awareness of the exceptions to these principles. Thirty-nine students enrolled in two beginning-level French courses at Emory University participated in the study.
The researchers chose eight structures that exemplified exceptions to grammatical patterns and randomly assigned them to the two classes; the assignment of treatments to the two classes was also random. The following semester the classes switched treatments so that all eight grammatical exceptions were taught using both techniques. The grammatical exceptions appeared in the same order they were presented in the textbook and subsequently in the classroom: (a) the contraction of \(de + le = du\) used as a preposition, (b) the contraction of \(a + le = au\) used as a preposition, (c) the use of the masculine possessive adjective before a feminine noun beginning with vowel, \(mon\), (d) \(ne\) (verb) \(pas\) \(de\), the replacement of \(un(e)\) by \(de\) after certain negated verbs, (e) \(dites\), the irregular form in the second person, formal \((vous)\), (f) \(cet\), the special form of the masculine demonstrative adjective used only before a masculine noun beginning with a vowel, (g) \(meilleur\), the irregular form of the comparative (in place of \(*\) \(plus\) \(bon\)), (h) -\(er\) imperative, the irregular form of the second person, familiar command. The aforementioned structures were taught approximately two weeks apart in the classroom.

In both classes, the researcher, Herron, followed the same methodology in her introduction of the grammar point by placing on the board various sentences with blank spaces. Students applied general grammatical rules to complete the blanks. The difference between the two treatments occurred when Herron introduced the exception. She wrote on the board a sentence requiring the application of an exception to the general grammatical pattern. The control group was taught the exception before being asked to apply the rule to the sentence. The experimental group received no explanation of the exception and was allowed to complete the sentence by responding orally to the teacher
who subsequently wrote the class’s answer on the board. After writing the incorrect response on the board, the teacher crossed it out and wrote over it the correct form accompanied by a brief explanation.

Tomasello and Herron (1988) administered a fill-in-the-blank test three times to assess the students’ progress in learning the exceptions. The majority of the test covered general grammatical patterns and was not exclusive to the targeted exceptions. In fact, the tests only included one example of each target structure exception. The first assessment came one to four days after instruction, the second was given 4 to 11 class days after instruction, and the third was administered for the midterm or final examination. After comparing the results of all eight target structures receiving the two treatments, the experimental Garden Path group demonstrated superior performance on all three tests.

Nevertheless, in explaining the cognitive effectiveness of the Garden Path Method, the researchers propose the following theory:

In the Garden Path approach, a discrepancy is created between a student’s own hypothesis (as manifested in the error s/he just produced) and what is grammatically correct. Since error correction occurs immediately after the student’s mistake, the Garden Path technique provides maximal opportunity for comparison of linguistic structures (Tomasello & Herron, 1991, p. 973).

The instruments in the current study employed the more accessible and widely-used term “grammar” in referring to issues of teaching a language’s forms and structures. The main foci for this study and the construction of the grammar-related items relate to Long and Robinson’s (1998) tripartite distinction between fonf, fonfs, and focus on
meaning as well as deductive and inductive approaches to grammar teaching (Tomasello & Herron, 1988, 1989, 1991). Primarily, the aim was to determine teachers’ and students’ opinions of how to approach grammar instruction—inductively or deductively, meaning-focused or grammar-focused—as well as the assessment of grammar.

ERROR CORRECTION

Closely related to instructed SLA, and grammar teaching in particular, is the area of corrective feedback. In many cases focus on form and grammar teaching attempt to influence students’ acquisition of structure and form preemptively while error correction and negative feedback specifically deal with learners’ errors after they occur. The questions included on this study’s instruments related to participants’ perceptions of the best way to deal with students’ errors, specifically in regard to issues of timing, degree of directness or explicitness, and nature of resulting explanations.

In addition to the Tomasello and Herron (1988, 1989) and White, Spada, and Ranta (1991) studies which explored the impact of different means of feedback, Carroll and Swain (1993), Lyster (1998), and Lyster and Ranta (1997) have all analyzed the role of feedback, what shape it should take, and the impact it has in enhancing second language acquisition.

In one of the most oft-cited studies into error correction Carroll and Swain (1993) studied the impact of various types of feedback on the acquisition of the English dative alternation by 100 adult, Spanish-speaking learners of English as a second language (ESL). In English some verbs allow the direct object, e.g., a package, to appear in one of two positions in relation to the indirect object, e.g., Ann, in dative verb constructions
such as *John sent Ann a package* and *John sent a package to Ann*. After administrating a pre-test to assess learners’ knowledge of the target structure and other grammatical structures, participants were divided into five groups according to the type of feedback received in individual feedback sessions. When participants in Group A made a mistake they were immediately told they were wrong and given explicit explanations as to how dative alternation works in English. Group B’s participants were simply told they were wrong with no accompanying explanation. Errors committed by Group C learners were reformulated correctly, i.e., recasts. Participants in Group D were asked whether they were sure their response was correct whenever they erred but given no explanation. Finally, Group Z’s learners were told nothing and received no intervention at all. Each group received 8 training items where they saw and heard two pairs of sentences demonstrating correct alternation and then they were asked to respond orally to two separate sentences and guess whether alternation was possible. During the feedback sessions the experimental groups received feedback on the ‘feedback’ items while no one received feedback during the testing, or ‘guessing’ items. This procedure was followed three times with two recall sessions following the initial experimental session.

Not surprisingly, the feedback groups significantly outperformed the control group (no feedback), with Group A (explicit, meta-linguistic) outperforming the other groups on both recall sessions. Only Group C’s performance on Recall session 1 did not result in a statistically significant difference from Group A. The researchers conclude from their results that explicit feedback does make a difference in adult SLA. Moreover,
they question claims that consciously learning about a language, as in acquiring declarative knowledge, does not aid in second language acquisition.

In their study, Lyster and Ranta (1997) analyzed the impact on students’ language production of six different feedback procedures used by four elementary-level French teachers in an immersion setting. After transcribing 18.3 hours of interaction from 4 classes, the researchers identified six methods of error correction used by the teachers including the following: 1) explicit correction—teacher indicates explicitly that an error was made, 2) recasts—teacher reformulates students’ utterances correctly, 3) clarification requests—teacher asks students’ to reformulate their utterance due to misunderstanding or error, 4) metalinguistic feedback—comments, questions, or suggestions made by the teacher referring to students’ utterances without explicitly stating the error or the correct form, 5) elicitation—teacher uses various techniques to elicit the correct form or word from students, e.g., beginning an utterance and asking students to finish it correctly, 6) repetition—teacher repeats incorrect form with adjusted intonation to highlight error.

The researchers wanted to determine to what extent, if at all, these corrective strategies directly affected students’ error awareness and production immediately following the corrective action. Several revealing findings resulted from their study. First, 38% of students’ errors received no corrective action from teachers. Of the 62% that received attention from teachers, 55% led to uptake, or students’ recognition of error by attempting to do something with the the teacher’s feedback. If there is no uptake there is topic continuation, in which case the teacher’s intention goes unheeded. Second, the most common strategy for dealing with students’ errors was recasts (55%); unfortunately,
recasts had the lowest probability of leading to uptake by learners’ (31%). Third, the most successful technique in achieving student uptake of errors was elicitation and this same strategy led to the highest percentage of student-initiated repairs (43%) as well as the highest rate of uptake by students (100%).

In a follow-up study using the same data set, Lyster (1998) set out to specifically analyze the nature of the 377 recasts used by teachers due to their overwhelming frequency and ineffectiveness in the previous study. He identified 4 types of recasts: 1) isolated declarative—correct reformulation with falling intonation and no additional information offered, 2) isolated interrogative—correct reformulation with rising intonation and no additional information, 3) incorporated declarative—correct reformulation incorporated into longer statement with additional information, 4) incorporated interrogative—correct reformulation incorporated into a declarative requesting additional information from the student. The tallies of each type demonstrated that isolated declaratives made up 67% and isolated interrogatives constituted 12% for a total of 79% of recasts neither offering nor requesting further information related to content from students.

Most of the teachers’ recasts were actually signs of approval of content and to a lesser degree well-formedness. The author observes the confusing nature of this sort of recasts for students:

with regard to approval, recasts have more in common with noncorrective repetition and topic-continuation moves than with other forms of corrective feedback. This reveals what must be a source of ambiguity for young L2 learners
as well as a dilemma for teachers whose mandate is to teach both language and content: namely, how to reinforce the substantive content of student messages while giving them clear messages about language form (p. 71).

These studies point to the effectiveness of making students’ aware of their errors by using various types of feedback and corrective strategies. As is the case with many studies in second language acquisition (SLA) the assessment of experimental interventions is performed fairly shortly after the experiment making it hard to draw conclusions regarding long-term acquisition. Studies in this area of research would do well to include more delayed testing and even longitudinal studies to determine the impact of different types of feedback on learners’ second language acquisition over the long term.

When learners make errors in their production, either spoken or written, L2 teachers must make several decisions. Several authors have cited Hendrickson’s (1978) well-known questions which have served to provide direction both for the study of error correction by researchers and the concrete implementation of error correction protocols by teachers in the classroom. These are 1) should learners’ errors be corrected?, 2) when should learners’ errors be corrected?, 3) which errors should be corrected?, 4) how should errors be corrected?, and 5) who should do the correcting? Teachers must answer these and other questions for themselves in order to successfully aid students in improving their foreign language proficiency. The items relative to error correction included in this study’s instrumentation principally focus on students’ and teachers’ opinions of direct
versus indirect means of addressing student errors to determine which approach each group feels is best for language learning.

TARGET LANGUAGE USE

Authentic, comprehensible input in the target language has been lauded by Krashen and Terrell (1983) as paramount in the acquisition of another language. The Natural Approach that these scholars have proposed with its theoretical base—the Monitor Model—posits that learners will only acquire a language if they are exposed to large amounts of the target language which is at a level of complexity and intelligibility just beyond their current level (i + 1). Comprehensible input, according to Krashen and Terrell, is fundamental in aiding students in their attempts to progress from just learning a language—as in conscious, declarative knowledge—to actually acquiring it—as in subconscious mastery akin to first language acquisition by children. Furthermore, Krashen and Terrell argue that students should not be forced to produce the foreign language until ready and that they should be allowed to use their native language when needed.

Similarly to Krashen (1982), VanPatten (1996) has argued that input, indeed, is crucial for second language acquisition to take place, but goes one step further in detailing from a psycholinguistic point of view how the processing of input facilitates second language acquisition. VanPatten’s Input Processing (IP) model is comprised of three principles: 1) learners process input for meaning before processing it for form, 2) if learners are to process non-meaning bearing forms, e.g., third person –s in English, they must be able to deal with informational, communicative content with little to no burden
on attentional resources, 3) most learners default to a ‘first noun’ strategy whereby they
assign the role of agent to the first noun or noun phrase they encounter. The IP model
makes a critical distinction between comprehension-based and processing-based
approaches to input. Since Krashen’s (1982) approach to input is primarily meaning, or
comprehension-based, it may not lead to any focus on form. VanPatten is much more
concerned with learners’ ability to control their attention to form-meaning relationships
during input processing as a means to expedite acquisition. While Krashen relegates the
acquisition of language forms to subconscious, acquisitional processes beyond learners’
control, VanPatten places a premium on cognitive, attentional processes that transpire
during input processing and how learners may make conscious moves to detect and
manipulate them.

In spite of these theoretical models emphasizing target language input, the
pervasive assumption that the target language should be the sole medium of
communication in the classroom has been questioned, more recently, by some scholars.
Rather than considering the native language as a crutch of sorts, Antón & DiCamilla
(1999) and Cook (2001) argue that the first language of students in the foreign language
classroom can be an excellent resource in facilitating FL language learning.

Alterations in the language used by foreign language teachers in order to promote
students’ comprehension seems appropriate to use in the classroom context. It would
appear useless for any teacher using any methodology to insist on language that is
completely unintelligible to the learner. On the other hand, it seems unproductive for the
teacher to use language too readily accessible to the learner in teaching the foreign
language, for example, using the students’ native language intermixed with sporadic words from the target language. Therefore, teachers’ use of the target language versus the students’ mother tongue in the foreign language classroom can be compared to a continuum ranging from high to low, high representing frequent use of the L2 to low representing infrequent use of the L2.

Duff and Polio (1990, 1994) have addressed the subject of L1, specifically English, and L2 use in the foreign language classroom and its possible causes. In their studies they set out to determine the amount and nature of target and native language use in the foreign language classroom. Their intent in the first study was primarily to quantify the amount of foreign language used in the classroom as compared to the L1, or English, and in their follow-up study they qualitatively categorized the use of English.

During the 1988-1989 school year at UCLA thirty-one foreign languages were taught, thirteen of which were included in their study. All of these classes were taught by native-speakers of each respective language, and each class had at least five students. All of the classes were second-quarter classes. Two fifty-minute segments of the class were observed and recorded by the researchers. An observation form was completed by the observers with basic information about the class. Students in each class filled out questionnaires soliciting information about their motivation to learn the target language, attitudes about the amount of English used in the classroom and amount of language students understood. Furthermore, each teacher was interviewed to ascertain his or her education and background, pedagogical philosophies, opinions concerning the use of English in the classroom, and departmental guidelines regarding English use.
In gathering the data, Duff and Polio (1990) used the following guidelines to quantify speech in the classroom every fifteen seconds:

L1: The utterance is completely in English

L1c: The utterance is in English with one word or phrase in the target language.

Mix: The utterance is, approximately, an equal mixture of English and the target language.

L2c: The utterance is in the target language with one word or phrase in English.

L2: The teacher’s utterance is completely in the target language.

Pause: No speech.

?): The utterance was not clear enough to be coded. (1990, p. 156).

In order to preserve the anonymity of the teachers, the researchers assigned letters to each language. After two observations the data showed a significant difference in the teachers’ use of the foreign language from class to class. The high was 100% in Language A’s classroom and the low was 9.5% in Language M’s classroom. The great disparity between classes becomes even more interesting due to the presence of observers. Apparently, the presence of the observers did not have a leveling effect wherein teachers aim for the perceived target of high L2 use.

Although reluctant to specify a causal relationship between the results and certain variables, Polio and Duff (1990) do identify some variables that could have influenced the outcome. First, the structural and lexical similarity of the target language to English
along with the departmental policy; second, the lesson content and the materials implemented in the lesson; finally, the amount of formal teacher training might have influenced the teachers in their language choice.

Four years after their initial study, Polio and Duff (1994) published a follow-up study that qualitatively examined the use of English in an attempt to determine when teachers tended to use it and for what purposes. In carrying out this study, the researchers chose to transcribe the language in the classroom from the previous audio-recordings. The six classes that used the greatest mix of the two languages were selected for transcription. After evaluating the classroom transcriptions the following categories were created to identify the functions of English use:

1. Classroom administrative vocabulary
2. Grammar instruction
3. Classroom management
4. Empathy/solidarity
5. Practicing English
6. Unknown vocabulary/translation
7. Lack of comprehension

Of the eight categories identified, the most surprising or non-intuitive seem to be “Empathy/solidarity,” “Interactive effect involving students’ use of English,” and “Practicing English.” Duff and Polio related “empathy/solidarity” with “interpersonal, rapport building purposes” (1994, p. 318). It seemed the teacher spoke English in order to
joke with the students and to show genuine concern or empathy. The “Interactive effect” referred to the tendency of the teachers to code-switch to English in reaction to students’ English use. As Polio and Duff (1994) argue, these cases of English use were not due to an apparent misunderstanding, but rather a simple acquiescence to the language choice of students. Given the teachers’ non-native fluency in English, the tendency to practice it does not appear surprising.

Another topic central to the larger issue of target language use in the classroom is the nature of the language used by the FL teacher and the modifications, or adaptations, performed to increase comprehensibility. This phenomenon of speech modification relates closely to the language typical of native speakers to non-native speakers, or ‘foreigner talk’. In first language acquisition several terms have been applied to this style of speech ranging from ‘motherese’ to ‘baby talk.’ Whether addressed to an infant or to an adult non-native speaker, the common element these pseudo-languages have in common is basic simplification with the ultimate goal being “comprehensible input” as prescribed by Krashen.

Ferguson (1971) was one of the first researchers to examine “foreigner talk.” Ferguson found three simplification processes used by native speakers when addressing non-native speakers. He identified omission, expansion, and replacement or rearrangement. Omission refers to the tendency of native speakers to exclude certain grammatical forms such as articles, inflectional morphemes, e.g. “*Boy eat candy, boy get sick.” Expansion alludes to the inclusion of forms not needed for a native speaker to understand, e.g. “You eat now!” Finally, replacement or rearrangement that includes
modifications in syntax such as “*You need towel”. Long (1996) summarizes foreigner talk by stating that it “tends to be well formed, delivered more slowly than speech to native speakers, with clearer articulation and fewer sandhi processes” and “employs shorter, syntactically or propositionally less complex utterances, and a narrower range of higher frequency vocabulary items” (pp. 416-417).

Similar to “foreigner talk” teacher talk in the foreign language classroom undergoes certain modifications in order to facilitate intelligibility. As Chaudron (1988) outlines, modifications in teacher talk can be categorized into phonology (speech rate, pauses, and prosody), lexis, syntax, and discourse (1988, p. 58). Chaudron provides an abundant amount of empirical research from various scholars backing his claims some of which will be reviewed below.

Studies from numerous researchers have shown the tendency of foreign language teachers to slow their speech when addressing non-native student speakers. Dahl (1981) conducted a study using six ESL teachers and six graduate students with no teaching experience. They were instructed to record a description task to four distinct audiences: native-speakers, advanced, intermediate, and beginning level ESL students. Dahl found that the rate of speech used for the native-speakers, advanced and intermediate ESL learners did not vary significantly. However, when the speech targeted the beginning learners the rate dropped dramatically. Likewise, Steyaert (1977) found that ESL teachers sped up their speech when telling a story to native-speakers versus their story telling to intermediate-level ESL learners.
Another tactic employed by teachers in the foreign language classroom in order to make themselves understood is the use of a much simpler, accessible vocabulary. Henzl (1979) reports that language selected by teachers and non-teachers when addressing L2 learners was more stylistically neutral, and included fewer idioms, more concrete and proper nouns, and fewer indefinite pronouns. Examples illustrative of this principle might be ‘man’ instead of ‘young guy’ or ‘Mrs. Smith’ instead of ‘the lady’.

Similarly, the syntax and discourse style of teachers addressing non-native speakers reflects the instructors’ tendency to facilitate learners’ understanding by excluding complex structures. Regarding discourse modifications, Chaudron (1988) mentions the inclination of teachers to repeat themselves frequently. In summarizing the adjustments made by proficient speakers of a language to lower-level non-native speakers Chaudron lists the following:

1-Rate of speech appears to be slower.
2-Pauses, which may be evidence of the speaker planning more, are possibly more frequent and longer.
3-Pronunciation tends to be exaggerated and simplified.
4-Vocabulary is more basic.
5-Degree of subordination is lower.
6-More declaratives and statements are used than questions.
7-Teachers may self-repeat more frequently (p. 85).

The items relative to target language use that appear on this study’s questionnaires pay particular attention to the following considerations: the use of English in the FL
classroom by the teacher, the teacher’s degree of proficiency in the target language, the
timing of the onset of target language use by the students, and the alteration of the
teacher’s language use for increased student comprehension. These issues lie at the heart
of practical, pedagogical decisions teachers must make in their use of the target language
classroom and were thus included on the questionnaires.

CULTURE

The connection between language and culture and the influence of each one on
the other has been a topic of debate for decades leading to the Sapir-Whorf, or Whorfian,
Hypothesis (Sapir, 1949). The radical version of this hypothesis, linguistic determinism,
claims that language determines thought while the weak version, linguistic relativity,
asserts that a specific language may influence perceptions of reality of those who speak
it. If indeed language and culture are so tightly interconnected then it would seem logical
that those who learn a foreign language as adults should also be exposed to the culture, or
cultures, of those who speak it. For this reason, some scholars have proposed that the
development of cultural knowledge and skills be included in foreign and second language
curricula alongside traditional linguistic skills—Kramsch (1993) and Lange (1999, 2003)
being two of the most noteworthy. Nelson Brooks (1968) argues convincingly that we
must take students beyond the mere presentation of cultural facts:

as long as we provide our students only with the facts of history or geography . . .
as long as we provide them only with a knowledge of the sophisticated structures
of society . . . or examples and appreciative comments on artistic creations . . . we
have not provided them with an intimate view of where life’s action is, where the individual and the social order come together, where self meets life (p. 212).

Similarly, Kramsch (1983) contends that in teaching culture teachers must integrate students’ affective and cognitive processes in interpreting cultural facts and maintains that students can successfully understand and interpret another culture’s attitudes and values only when they are aware of their own. Kramsch encourages teachers to broaden their knowledge of not only the language but the patterns of usage. Language teachers who experience success in teaching culture must become familiar with the way native speakers might think and react by gaining an increased familiarity of their own.

In her analysis of eight, first-year German language textbooks Kramsch (1987) found an abundant use of introductory dialogues with little to no socio-cultural explanation or context attached. Students are made to repeat statements of fact uncritically without discussing their cultural truth. An example from one of the textbooks helps to illustrate this point. In presenting the competitive nature of sports a comparison is made between Germany and the United States in table form with lists presented side-by-side. On the United States’ side “true amateurs” and “self-supported” are listed opposite Germany’s “state amateurs” and “state-supported”. The German list contains several characteristics for which the American list simply lists a “?” in parentheses implying that there is no American counterpart, i.e., origins: national socialism, (origins: ?); sports = politics, (sports = ?); financed by nationalized industry, (financed by ?). Kramsch points out that the textbook leads students to infer opposing ideals for German sports to those which undergird American ideology: freedom, free enterprise, democracy, sports as an
apolitical activity. The textbooks minimized cultural differences between the target and the native cultures and appeared to lead many students to believe that different cultures and societies are not, in effect, different worlds, rather similar worlds with distinct labels (Kramsch, 1987).

In his textbook, *Teaching Culture*, Seelye (1997) proposes that the key to understanding what people say is context and that context can serve as a powerful organizer of experience. From a pedagogical standpoint, Seelye defines and explains how to integrate concrete exercises like culture assimilators, culture capsules, and culture clusters. He also includes 6 practical, instructional goals for the teaching of culture in the language classroom which are presented below:

Goal 1—Interest: students demonstrate curiosity about another culture.

Goal 2—Who: students recognize the crucial impact of role expectations and the way in which other social variables such as age, sex, social class, religion, and ethnicity affect how others speak and behave.

Goal 3—What: students realize that effective communication hinges on the discovery of culturally conditioned images that are evoked in the minds of people when they think, act, and react to the world.

Goal 4—Where and When: students recognize that situational variables and convention shape behavior in important ways.

Goal 5—Why: students understand that human behavior is constrained by options made available by each society to satisfy basic physical and psychological needs, and that cultural patterns are interrelated and support need satisfaction.
Goal 6—Exploration: students can evaluate a generalization about the target culture in terms of corroborating evidence, and have the skills to locate and organize information about the target culture from other sources, i.e., library, mass media.

As demonstrated above, the principal argument made by proponents of culture teaching is that language cannot be taught without reference to the culture(s) of those who speak it, and that the cultural material presented must go far beyond de-contextualized facts that do not shed light on people’s attitudes, motives, and world view. Many cultural and pragmatic issues arise when learning a language spoken by people of a different society who may espouse a different world view, e.g., the tú vs. usted distinction in Spanish or the honorific system in Japanese. The importance of the connection between foreign language pedagogy and culture is further emphasized by the central role culture plays in the Standards for Foreign Language Learning in the 21st Century which are reviewed in the standards section below. As the world becomes more interconnected and interdependent economically and socially and as more language learners acquiring second languages interact directly with native speakers, the role of culture teaching in language classrooms becomes even more crucial.

Even with all that has been written in favor of culture teaching in the FL classroom and the crucial role of culture in learning a foreign language more generally, the increasingly popularity of language classes for specific purposes appears to reflect the desire by some to focus more strictly on language learning to the exclusion of culture learning. In these language courses the objectives are to cater language learning to specific contexts and circumstances so adult learners may achieve greater proficiency in
that environment. Ever more common are FL classes which aim to instruct learners, for example, in ‘Business Spanish’ or ‘Spanish for Medical Purposes’. The term *English for Specific Purposes* has become quite common in the field of English language learning in referring to curricula where the primary goal is to prepare students to function in a specific communicative context with the individual’s needs as essential while not overly stressing culture. As societies become more global, and life becomes more fast-paced, many adult learners have very limited time and very particular goals with language learning. Thus, they prefer to be instructed in the specific domain that they personally will need to master for professional or personal reasons, e.g., business English, English for airline pilots, medical Spanish, Academic English., rather than spend time on what some may feel are peripheral cultural facts.

Two of the questionnaire’s items concerning culture are more global than many of the other items and were intended to determine participants’ overall reactions to, 1) the amount of culture to be taught in the FL classroom as compared to language skills, and, 2) the amount of cultural knowledge an effective teacher should possess. The third culture question relates more specifically to the use of authentic, cultural materials in the teaching of the foreign language and whether participants feel they are necessary to effectively teach culture. As the place of culture in foreign language teaching receives ever increasing emphasis from national organizations and the standards they produce, the goal of these items was to see if students’ and teachers’ perspectives paralleled not only each others’ but the aforementioned standards.
COMPUTER-ASSISTED LANGUAGE LEARNING

The use of computers and their effectiveness in facilitating second language learning has been explored by many scholars over the last two decades (Beauvois, 1998; Blake, 1998, 2001; Bush & Terry, 1997; Kern, 1995b; Kern & Warschauer, 2000; Muyskens, 1997; Payne & Whitney, 2002). Computers have allowed students to interact with authentic language and culture visually through video, textual, and graphical media as well as aurally through audio media. Computer Assisted Language Learning (CALL) software programs enable students to engage in extensive practice with grammar, vocabulary, and aural comprehension outside of classroom at their own pace with meaningful feedback provided immediately. These software programs when delivered in a CD-ROM presentation or uploaded to the internet allow the teacher to assign students activities to be completed outside of class.

More recent integrations of the computer in the second and foreign language classrooms have taken advantage of the World Wide Web (WWW) and Local Area Networks (LAN) to foster meaningful communication. Computer-mediated Communication (CMC) has presented itself as a viable way to help language learners improve their language skills. The nature of CMC is such that, in general, language students have the opportunity to communicate more with more people of a more diverse linguistic and cultural background more freely. The research so far has corroborated the conclusion that CMC affords a greater percentage of L2 students the opportunity to interact in the L2 by not only re-connecting humans to each other but allowing for an initial connection for more reserved students (Kern, 1995b; Warschauer, 1996).
As alluded to above, one area of CMC research that proved most fruitful and conclusive in its results was the quantitative analysis of student-to-student interaction patterns in CMC versus face-to-face discussions. Students seemed to interact much more when communicating via the computer both at the class and the group level. In Kern’s (1995b) study 40 students in two French 2 classes discussed the same topics in a face-to-face and a virtual discussion format. The CMC discussion preceded the oral discussion and both were directed by the teacher with students responding to each other’s comments at the class level. The researchers chose to focus on the fifth electronic discussion that took place in both sections about mid-semester and further decided to videotape, audiotape, and transcribe the oral discussions. The author concedes that inferential statistics would not be appropriate for such a small sample size; however, the resulting descriptive statistics are quite telling. On average, across both sections, the CMC discussion resulted in two to three and a half more turns per student as compared to the oral equivalent. Furthermore, in the CMC exchange all students participated while in the oral interaction almost a third (4 of 14) in Section 2 did not participate at all. Another revealing finding was that students produced two to four times more sentences on the computer than spoken T-units (the oral equivalent of a sentence). Finally, in Section 1 students directly addressed other specific students over six times as often in the CMC condition than in the oral condition. Even more discrepancy was found in Section 2 where 232 sentences were used to address specific students in the CMC medium as compared to only one in the spoken conversation. These last two findings appear logical
as paralinguistic factors such as eye gaze contribute heavily to turn-taking, gaining the floor, and directly addressing others without the explicit use of names.

The resounding results are tempered, but not altogether dismissed, by certain limitations mentioned in the article. First, the author makes reference to the small sample size of students and the fact that the oral discussions ended up being 1-2 minutes shorter than the CMC exchanges. Kern (1995b) also points to the unequal participation pattern of one of the instructors between the oral and electronic discussion, the latter receiving less intervention on the instructor’s part. Third, the discussion order—CMC first and oral second—may have contributed to students feeling like the topic had been exhausted and that the discussion had become redundant. In spite of these design weaknesses it seems hardly feasible that such limitations could explain away the vast quantitative difference in participation patterns between the oral and electronic discussion formats. Indeed, participation appeared to be more evenly distributed and more accessible to a higher percentage of students in the CMC condition.

Similar to Kern’s (1995b) finding, a more recent study by Liu and Sadler (2003) found that peer groups using CMC produced more comments than traditional face-to-face groups. Kern’s study demonstrated that CMC discussion groups availed themselves effectively of a wide array of discourse functions other than purely managerial language, such as asserting personal opinion and asking for opinions from others. Although Liu and Sadler’s study involved very different students, international ESL students, in a different task, peer review of writing, than that of Kern, their findings were much less positive with respect to the effectiveness of CMC for instigating concrete revisions. They found
that face-to-face groups made fewer comments on writing but that a higher percentage of comments received attention in subsequent revisions unlike revisions made through CMC. The learners in Liu and Sadler’s study appeared to prefer face-to-face interactions in synchronous peer reviews and appreciated the ability to use body language as an additional help in understanding and interpreting their partner’s comments.

One weakness of Kern’s (1995b) study that Warschauer (1996) addresses in his research design is the limiting of face-to-face and CMC comparisons to whole-class discussions. Warschauer appropriately acknowledges that discussions at the class level are particularly intimidating for shy students. Whether or not CMC participation rates would continue to surpass those of face-to-face participation in small group discussions motivated Warschauer’s analysis of 16 ESL students’ participation patterns in both mediums. The 16 students were divided into four groups of four students each and were asked to discuss questions relevant to their curriculum for 15 minutes in each condition, i.e. electronic and face-to-face. In three of the four groups the results demonstrated that electronic discussions proved twice as equal in participation distribution than did the face-to-face discussions. One possible explanation offered by the author is the fact that the three groups impacted by CMC had Japanese students while the one group apparently unaffected by CMC had no Japanese students. It was the Japanese students who hardly participated at all in the oral discussions. Warschauer raises the important issue of culture, a construct closely related to personality, which spurs the question: Does the equalizing influence of CMC on participation rates affect different personality types?
differently? Are introverted, shy students the only ones who benefit from avoiding the additional emotional and psychological burden of face-to-face interaction?

Less conclusive are the results of studies that have examined the relationship between the development of oral ability via CMC. Beauvois (1998) analyzed the oral achievement of two groups of fourth-semester French students three times over the period of one semester. The control group and the experimental groups differed in only one key aspect: once a week the experimental group discussed the topics from the text electronically while the control group discussed them orally. A statistical analysis of the average means between the two groups on the oral exams proved significant at a level of \( p<.05 \). Statistically speaking the results of the experiment lead one to believe that it was not by chance that the experimental group surpassed the control group.

In spite of Beauvois’s efforts to include a list of precautions that were taken to ensure the control of intervening variables, two key design factors were apparently overlooked by Beauvois (1998) making her findings quite suspect. First, the assumption that randomly assigning 83 students already enrolled in French 212 to four sections resulted in groups which were not significantly different. Without a pretest, posttest design there is no way to tell if the control group and the experimental groups represented different populations before the treatment, and, therefore, whether the treatment created the significant difference in the two populations after its administration. Second, an assessment instrument comprised of a 10-minute oral exam with one’s own instructor without calculating reliability, inter-rater or intra-rater, does not instill confidence in its validity. The author points out that her study is merely a pilot study but even as a pilot
study the conclusions drawn from questionable assessment practices can be, at best, questionable.

In contrast to the study by Beauvois (1998), Payne & Whitney (2002) conducted a study based on solid research design which allows their results to be more insightful and robust. This study implemented a pretest, posttest quasi-experimental design with two experimental groups and two control groups. A modified SOPI was administered by examiners who were not the teachers of the participants. The pretest inter-rater reliability was .86 and the posttest reliability was .94. An ANCOVA test was run with the pretest as a covariate in order to take into account the students’ level before the treatment. The authors included this statistical analysis after taking note that the groups were significantly different prior to the administration of the treatment, a crucial factor not accounted for in Beauvois’s study. The results of the ANCOVA showed that the experimental groups as a whole outperformed their control group counterparts significantly at the p<.05 level after taking into consideration the pretest. With such a solid research design backing them the authors claim that “these findings suggest that the participants spending half of their instructional time in a synchronous online environment were advantaged in their oral proficiency development over those meeting face to face in the classroom” (Payne & Whitney, 2002, p. 20).

The findings from a recent Ph.D. dissertation further confuse the issue of how CMC and L2 oral ability relate. Heather (2003) investigated the possibility of using a CMC test in place of an oral exam. He queries, “Can one make inferences about a student’s speaking ability based on their performance on a computer-mediated (i.e.
Students from two third-semester French classes participated in the study. Both classes had a CMC component as part of the regular curriculum during the semester. All students were tested in the same groups of three for both the CMC test and the face-to-face test using two different topics. After addressing the complex issue of creating an adequate rubric and assessment instrument that would be applicable to both CMC and speech, two independent raters assessed the students’ language. The results from Heather’s study indicated that there were no significant correlations between performance on the CMC test and the group oral exam. Similar to other studies in the CMC tradition, Heather’s comparison of face-to-face and oral exams included quantity of language produced, lexical density, type/token ratios, and language function. However, the major contribution of this study is the finding that an assessment of one’s ability on a CMC-based exam does not accurately predict performance on an oral exam. Like most researchers, Heather experienced many difficulties in conducting the experiment, not the least of which were a very low inter-rater reliability and the arduous task of creating a satisfactory rubric and assessment instrument to suit both electronic and face-to-face interactions.

Due to space and time constraints on the present study’s questionnaires and allotted class time for administration, only one item on the questionnaire is dedicated to the use of computer-based technologies. On the effective foreign language teacher questionnaire the item includes the adverb “frequently” as an attempt to better elicit participants’ opinions of the use of technology in the classroom. Some feel, as Blake has argued (2001), that although teachers will most likely not be replaced by computers,
teachers who use computers will be replaced by those who don’t. Nonetheless, what are the opinions of this study’s participants regarding the use of computer-based technologies in the FL classroom?

COMMUNICATIVE APPROACHES TO LANGUAGE TEACHING

In reaction to the Chomskyian focus on Universal Grammar and the competence/performance dichotomy, Hymes (1972) argued convincingly of the need to include the idea of ‘communicative competence’ in any description of language acquisition. Speakers do not acquire and speak language in a vacuum and must know when to say what to whom. A few years later Canale and Swain (1980) proposed four aspects of communicative competence: 1) grammatical competence, 2) sociolinguistic competence, 3) discourse competence, and, 4) strategic competence. Grammatical competence refers solely to the ability to manipulate the linguistic code and pertinent structures to transmit meaning. Sociolinguistic competence is synonymous with the capacity to take into account contextual factors such as topic, setting, and audience. A speaker who demonstrates sociolinguistic competence knows how to produce appropriate language and interprets language appropriately given diverse contexts. Discourse competence alludes to the combining of grammatical structures and lexical items to produce a cohesive, unified spoken or written text within different genres. Finally, strategic competence has reference to the speaker’s ability to integrate communicative strategies, both verbal and non-verbal, to understand language and make oneself understood when communication has broken down or simply to enhance what is being said.
Bachman (1990) termed his more recent instantiation of communicative competence as *language competence*. His model can be divided into two categories that branch off from language competence: *organizational* and *pragmatic* competence.

*Organizational competence* refers to the ability to use formal properties of the language to combine into a coherent text, spoken or written. This competence is further divided into *grammatical competence* and *textual competence*. Bachman divides pragmatic competence, or the sensitivity to context and audience, into *illocutionary competence* and *sociolinguistic competence*. Illocutionary competence relates to speech act theory and pertains to speakers’ ability to determine the intent or function of an utterance regardless of the grammatical form that utterance takes. Sociolinguistic competence, then, refers to the use of certain language conventions registers that would be considered appropriate given various contexts, e.g., dialects, interlocutors, location.

The birth of ‘communicative competence’ as a viable construct to be considered as one aspect in designing L2 curricula and the inability of the Audio-Lingual Method to produce competent users of the target language influenced the appearance of communicative language teaching by causing scholars and teachers to rethink fundamental goals and objectives with language learning and teaching. Communicative language teaching became a viable L2 methodology in foreign and second language teaching and represented much more than just teaching students about sociolinguistic competence as reflected in most models of communicative competence. Savignon (1972) revolutionized language teaching and its fundamental objectives with her findings concerning communicative language teaching. In spite of its age, Savignon’s landmark
study in communicative teaching stands as very convincing evidence in favor of a communicative approach for the improvement of fluency. Her study consisted of 42 students enrolled in three sections of beginning French at the University of Illinois at Urbana-Champaign for an 18-week semester. All three groups met four times a week for 50-minute class sessions and an additional 50-minute session spent in a French-related activity. The control group met in a language laboratory practicing the material presented in the class. The two experimental groups were divided further into two groups of six or seven and received alternative instruction. One set of groups received instruction in English about French culture, cuisine, politics, films, and even encounters with French students. The other experimental groups received training in performing specific communicative acts that progressed from simple greetings to making a date, asking for directions, and even discussing current events.

The experimental design controlled for variables such as language aptitude, verbal intelligence, high school percentile rank, and previous exposure to French—any student who had received exposure to French prior to the semester was excluded from the study. None of these independent variables showed a significant interaction effect with the dependent variable. An attitude questionnaire was administered three times throughout the semester as were tests of linguistic and communicative competence at the end of the semester. Savignon’s (1972) study produced a wealth of interesting data, foremost of which pertains to communicative ability. Students in the communicative experimental group significantly outperformed the other groups, both experimental and control.
The implications for improving fluency and communicative ability from Savignon’s (1972) study were quite clear, but what about accuracy? Students’ performance on standardized proficiency tests of reading and listening showed no significant difference as did students’ final grades. Savignon minimizes these results by stating that they were “evidence of the need to distinguish between communicative competence on the one hand and linguistic competence on the other, noting that it is the latter which is typically rewarded in the FL classroom” (p. 17).

Communicative language teaching served as the impetus behind several other approaches to language learning all based on the assumption that meaningful, contextualized, and realistic communication can lead to improved L2 proficiency. The concrete pedagogical approaches of communicative language teaching that receive attention on this study’s questionnaires include Total Physical Response (TPR), interaction and negotiation of meaning, and task-based teaching. Each approach with supporting theoretical explanations will be outlined below.

Before individual learning styles began to receive much attention in foreign language learning and teaching, Asher’s TPR (1966) approach proposed that language learning might be facilitated through kinesthetic means, or a total physical response, by using commands with students. Asher (1972) bases his approach to language learning on the assumption that children’s first language acquisition can serve as a model for adult language acquisition. He asserts that children begin acquiring their first language primarily by responding physically to parents’ commands before being able to produce coherent, spoken language. Parents and other care-givers generally use commands that
require children to execute movements with their body as in “Come here,” or “Give the toy back to Johnny.” Similarly, Asher contends, adult learners of a second language master listening far before speaking. Asher summarizes the rationale behind TPR in the following manner, “The particular sequence is listening before speaking and the mode is to synchronize language with the individual’s body” (p. 134).

In an early review of studies into the effectiveness of TPR in teaching listening, Asher (1966) documents the overwhelming success of experimental groups in outperforming control groups on listening comprehension tasks. These studies included both Japanese and Russian as the target languages and involved college students as well as middle and elementary school students. A study conducted in German classes demonstrated that over a longer training period as compared to earlier studies, beginning students in the experimental group performed significantly better than students in first and second semester German who had at least 35% more amount of exposure to German. An assessment of reading was administered to both experimental and control groups with the expectation that the control groups would score much higher than their experimental counterparts given their increased exposure to reading and writing. The statistical analyses did not result in a significant difference between groups, meaning that the control groups which had been exposed to reading and writing did not outperform at a statistically significant level those who hadn’t. Later, Asher, Kusudo & De la Torre (1974) tried out the TPR technique in the Spanish foreign language classroom and published another article, Learning a Second Language through Commands: The Second Field Test, which demonstrated, again, the noticeable impact of TPR on learning gains in
listening comprehension as compared to other methods. After 45 hours of training with
the TPR method, a group of college students outperformed a group of high school
students with 75 hours of training and another group of college students with 150 hours
of training. As a result of the technique’s success, Asher (2003) has published multiple
editions of a practical guidebook for teachers helping them to integrate his TPR approach
in the classroom. It is interesting to note that Asher’s theoretical premise for his
convincing success with TPR—behaviorism—has since been shown to be faulty.
Although this theoretical shift does not negate his findings, it does have fascinating
implications for future theory building in SLA, specifically listening comprehension.

Gass & Varonis (1994), Pica (1988, 1994) and Long (1996) have explored the
relationship to second language acquisition of interaction and negotiation of meaning
between native and non-native speakers and two non-native speakers. Researchers
interested in this field investigate the impact of different types of communicative
exercises on the nature and quantity of the target language exchanged during these
negotiations of meaning. Another variable of interest in this research tradition are the
identities of the interlocutors and how they might influence the amount and quality of
interaction. Gass summarizes the principal notion of interaction and negotiation of
meaning in second language acquisition this way: “conversational interaction in a second
language forms the basis for the development of language rather than being only a forum
for practice of specific language features” (Gass, 2003). Long (1996) has articulated a
very coherent and precise explanation of the Interaction Hypothesis and what it proposes:
negotiation for meaning, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways (p. 451-452).

Furthermore:

it is proposed that environmental contributions to acquisition are mediated by selective attention and the learner’s developing L2 processing capacity, and that these resources are brought together most usefully, although not exclusively, during negotiation for meaning. Negative feedback obtained during negotiation work or elsewhere may be facilitative of L2 development, at least for vocabulary, morphology, and language-specific syntax, and essential for learning certain specifiable L1-L2 contrasts (p. 414).

Just as the development of communicative competence among students represents one pedagogical strand of Communicative Language Teaching, task-based instruction represents a concrete pedagogical instantiation of interaction theory and is another strand of communicative teaching. Skehan (1996, 1998) has been of the most prominent scholars to examine and justify task-based language teaching wherein students are taught language through the execution of meaningful, concrete and purposeful tasks whose successful completion is measurable. The assumption behind a task-based approach to L2 teaching is that when students engage in completing tasks they are activating naturalistic language acquisition mechanisms which cause the interlanguage system to be stretched
and overall development to progress (Skehan, 1998). Skehan has offered the following five characteristics of tasks within a task-based teaching curriculum:

1) meaning is primary;
2) there is some communication problem to solve;
3) there is some sort of relationship to comparable real-world activities;
4) task completion has some priority;
5) the assessment of tasks is in terms of outcome (p. 95).

Skehan (1996) defines a ‘task’ in language teaching as “an activity in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome” (p. 38). While many activities in the classroom incite meaningful communication, tasks engage students in scenarios that carry meaning beyond the classroom (Skehan, 1996). Long and Crookes (1992) emphasize that a successful task-based syllabus must reflect the types of tasks learners will encounter as indicated by a needs assessment.

One of the strengths of task-based learning, i.e. motivated negotiation of meaning, may actually result in being detrimental to learner’s linguistic development. This concern regarding tasks relates to students’ attentional resources. In a task-based approach students have something worthwhile to communicate and, thus, content may become all-encompassing at the expense of accuracy. That is, if students are so focused on completing a task and communicating they will sap their available cognitive and attentional capacity which could lead to very inaccurate output. In order to prevent such attentional overload in completing tasks, Skehan (1998) proposes three essential
considerations that should govern their construction, selection, and integration in the classroom: 1) code complexity, 2) cognitive complexity, and 3) communicative stress. Code complexity refers to the difficulty of the language needed to complete the task, grammatically and lexically. Cognitive complexity includes cognitive familiarity and cognitive processing; the former refers to one’s background with the topic and discourse genre while the latter refers to the load imposed on students’ processing ability. Closely related to processing difficulties is the final issue Skehan points to, which is communicative stress. Communicative stress relates primarily to time limits and external pressures under which the task is to be completed. These three issues stem naturally from Skehan’s required characteristics of tasks, namely, that they be similar to real-life and that they be meaning bearing.

Research into task-based instruction has been difficult to perform given the multifaceted nature of ‘tasks.’ As Ellis (1994) points out, most studies have focused on specific variables and their effect on subsequent interaction. One significant strain of research has investigated the effectiveness of one-way interactions, such as story telling and direction giving, versus two-way interactions, such as games and exercises where each student holds part of the information needed to complete the task. These two types of tasks differ in that the former does not necessarily require language exchange while in the latter case language exchange is obligatory. Studies by Long (1980), Doughty and Pica (1986), and Newton (1991) showed that two-way interactions fostered more linguistic interaction than one-way tasks. Other task-based variables that have been
investigated are task familiarity, task participants, task goals, task organization (Skehan, 1998). Skehan summarizes research into tasks in this manner:

One could generalize and claim that, other things being equal, tasks are more effective if they contain information distributed amongst different participants, so that interaction is more likely, and may become of better quality (p. 114)

Communicative language teaching, at heart, encourages students to primarily produce language in meaningful contexts, either spoken or written, and tends to de-emphasize preoccupation with grammar. A fundamental SLA theory that underpins this approach is Swain and Lapkin’s Comprehensible Output Hypothesis (1995). While Krashen has emphasized the role of input in achieving grammatical accuracy, Swain (1995) proposes three ways in which output facilitates accuracy. First, output may require students to ‘notice’ a gap between what they are capable of producing and what they would like to produce. Second, language production serves as a way to test hypotheses about comprehensibility and linguistic well-formedness. Third, output may be used as a metalinguistic tool through which learners can control and internalize linguistic knowledge. Swain (1995) summarizes the proposal of the Output Hypothesis:

In producing the L2, a learner will on occasion become aware of (i.e. notice) a linguistic problem (brought to his/her attention either by external feedback (e.g. clarification requests) or internal feedback). Noticing a problem ‘pushes’ the learner to modify his/her output. In doing so, the learner may sometimes be forced into a more syntactic processing mode than might occur in comprehension” (pp. 372-373).
Swain (1995) hypothesizes that the cognitive task of formulating ‘pushed’ output causes learners to move from a semantic processing level characteristic of comprehension to a syntactic level required of production.

In a study of 18 eighth-grade students enrolled in an early French immersion class, Swain and Lapkin (1995) formulate the following research questions: 1) Do young adolescent learners ever become aware of gaps in their linguistic knowledge as they produce L2?; 2) When, and if, they do recognize gaps, what do they do?; 3) As part of their linguistic problem-solving process, do these learning engage in grammatical/syntactic analyses?

Researchers gave the students a writing task which included a short, one to two paragraph composition on a specific, familiar topic. The students met with the researcher in a small, quiet room and were instructed to verbalize their writing process (a think-aloud protocol) while drafting their first version and again during their revisions. The researcher demonstrated how this was to be done by working through a mathematical problem aloud. If the student paused for very long, the researcher would prompt the student with “What are you thinking?” and in no case did the researcher provide any help, linguistically or otherwise.

The think-aloud sessions were recorded, transcribed, and coded. The researchers limited their data collection to episodes that were language-related. To ensure reliability, four researchers independently categorized all language-related comments into the following seven categories: 1) sounds right/doesn’t sound right, 2) makes more sense/
doesn’t make sense, 3) applied a grammatical rule, 4) lexical search, 5) translation, 6) stylistic, 7) spelling.

The authors found that the participating students indeed found gaps in their linguistic knowledge and engaged in thought processes which play a role in learning a second language. Swain and Lapkin (1995) point out, however, that many of the students’ comments represented faulty notions about the L2 and incorrect generalizations were made. They conclude that along with opportunities for output, appropriate feedback may be a needed catalyst in expediting the acquisition process. Specifically, they state, “It is important to reiterate that output is not the only source of second language learning” (Swain & Lapkin, 1995, p. 386).

These communicative approaches to language learning encourage students to communicate more frequently and more meaningfully. Therefore, the communicative classroom is often student-centered where group and pair work is encouraged to maximize students’ opportunities to produce language. The issues of communicative language teaching that are addressed on the questionnaires relate to group and pair work, task-based teaching, Total Physical Response, exchange of meaningful information versus grammatical information inside and outside the classroom, and assessment of language interaction. All of these concerns in communicative language teaching revolve around the preeminence of meaningful communication as compared to detailed, detached grammatical analysis.
ASSESSMENT

The crucial role of language assessment practices in classroom language learning has been discussed extensively by Shohamy (1997, 1998) and Bachman (1990, 1996). Assessment practices in language classrooms reveal what teachers and administrators feel it means to know a language and which is the best way to determine the extent of that knowledge. Discussions of language testing revolve around fundamental issues such as the validity and reliability of assessment instruments, as well as the subsequent washback effect on curricula and teaching practices. Do language tests appropriately evaluate the knowledge base and skills set language students have learned in a way that allows them to demonstrate those skills and knowledge to their teachers?

Shohamy (1997) has analyzed closely the impact of test method on students’ outcomes as well as the consequences and uses of language tests. In reviewing research on the effect of testing method on students’ scores, Shohamy concludes that the chosen method of testing clearly affects the resulting scores of test takers adding that “test methods in a variety of language skills may be considered unfair unless they can be shown to be relevant to the tests construct” (p. 343). In an earlier article, Shohamy (1984) reports on a study she conducted that examined the effect of various methods used for testing reading comprehension. Her study revealed that multiple choice questions were easier than open-ended ones on two separate texts and that low-level students were affected more significantly than higher-level students. A very telling finding was the discovery that the language in which the reading questions were presented made a difference on students’ performance. That is to say, students performed better when the
questions asking them what they had read in the target language were posed in their native language—an issue taken up on the current study’s questionnaire.

Regarding her inquiry into the uses and consequences of language tests, Shohamy uncovered a wide range of both declared, or explicit objectives, and undeclared objectives. In some cases tests were used to alter teachers’ curricular focus, i.e., washback, while in other instances, tests were strictly used for bureaucratic or political purposes such as raising the prestige of a specific language. Other unethical uses of tests by central, often times governmental, agencies include the desire to manipulate educational systems, to control curricula, to define knowledge, and to communicate educational agendas. The author concludes by pleading for more democratic uses of language tests where all stakeholders participate in the decision-making process. The title of a more recent chapter on language assessment written by Shohamy (1998) seems to sum up her work and recommendations for language learning assessment: *Evaluation of Learning Outcomes in Second Language Acquisition: A Multiplism Perspective*. Multiple sources of language evaluation via multiple testing methods present the most valid and reliable approach to assessing second, or foreign, language proficiency.

One pressing issue in language testing that Bachman (1990, 1996) has addressed relates to the relationship between a language test and real-life language tasks. Therefore, the driving force behind the design, development, and use of any language test, according to Bachman (1996), should be the degree of correspondence between language test performance and language performance in non-test situations. In order to ensure a close correspondence between the two, test designers must pay close attention to the
characteristics of the language use situation and tasks and the language test situation and tasks. Additionally, Bachman suggests that the characteristics of individuals as language users closely correspond with test takers’ characteristics, specifically, language ability, topical knowledge, and affective schemata. The increased emphasis on authenticity in language testing has led to the use of various terms to define such real-life tests, namely, pragmatic, communicative, performance, and authentic.

The assessment items included on the instruments participants completed for this study cover issues relative to the use of English as a valid means of demonstrating mastery of listening and reading as well as the validity of communicative testing strategies with productive skills (i.e., speaking and writing) with its tendency to encourage negotiation of meaning and to de-emphasize strict grammatical accuracy. In short, participants were asked to express their perspective on the intersection between language assessment and communicative approaches to language teaching.

Although the literature review for each category provided here is far from exhaustive, it provides sufficient background to highlight the current relevance and to justify the inclusion of items from each category on the study’s instrumentation. A complete list of relevant references for each item on the questionnaires will be presented in table form in Chapter 3 in the discussion of instrument design and item construction.

SECOND LANGUAGE LEARNING AND TEACHING STANDARDS

In spite of the difficulty in specifying which specific teacher behaviors result in concrete student learning gains, a movement has arisen in the L2 teaching profession to establish concrete standards to which L2 teachers and L2 learners may aspire in their
quest for effective teaching and learning in the classroom. A monumental collaborative effort to draft standards for foreign language learning was published in 1999 with the title *Standards for Foreign Language Learning in the 21st Century*. Some of the professional organizations who collaborated in the drafting of these standards include the American Council on the Teaching of Foreign Languages (ACTFL) and eight language-specific organizations representing French, German, Spanish and Portuguese, Russian, Classics, Chinese, Japanese, and Italian. Within the field of second language learning and teaching, the association of Teachers of English to Speakers of Other Languages (TESOL) has also put together specific standards for language learning and teaching. These different sets of standards have been carefully articulated by experts to include central concerns in the teaching and learning of languages, and, therefore, help establish instructional goals for teachers. Even before the first version of the national standards were done being published in 1996, “the impact was being felt in local states and local districts . . . some states have initiated assessments tied to selected standards, and growth in elementary programs can be directly tied to the earlier starts advocated in the standards” (p. 15). The standards, in effect, have had a homogenizing effect on foreign language learning and, as I argue below, to a certain extent on foreign language teaching. Hence, the standards reviewed below help to further illuminate the discussion of current thinking regarding effective L2 teaching.

While the *Standards for Foreign Language Learning in the 21st Century* do not necessarily constitute specific curricular or sequential recommendations, they do make
the connection between the standards and improved teaching and learning as illustrated in
the statement below taken from standards published in 1999:

While they reflect the best instructional practice [emphasis mine], they do not
describe what is being attained by the majority of foreign language students. The standards described within these pages will not be achieved overnight; rather, they provide a gauge against which to measure improvement [emphasis mine] in foreign language education in the years to come (p. 28).

The standards are comprised of five general goal areas which shape their prescribed guidelines of what learners should know and be able to do. These are communication, cultures, connections, comparisons, and communities. Each category is comprised of two to three concisely stated standards with the ultimate goal for students being ‘Knowing how, when, and why to say what to whom.’ The standards as outlined reflect a desire and need to take the language outside of the classroom by not only addressing linguistic and pedagogical issues, but also cultural, societal and interdisciplinary concerns. Each of the five C’s with the corresponding set of standards is reproduced below as taken from the published volume.

**Communication:** Communicate in languages other than English.

Standard 1.1: Students engage in conversations, provide and obtain information, express feelings and emotions, and exchange opinions.

Standard 1.2: Students understand and interpret written and spoken language on a variety of topics.
Standard 1.3: Students present information, concepts, and ideas to an audience of listeners or readers on a variety of topics.

**Cultures:** Gain knowledge and understanding of other cultures.

Standard 2.1: Students demonstrate an understanding of the relationship between the practices and perspectives of the culture studied.

Standard 2.2: Students demonstrate an understanding of the relationship between the products and perspectives of the culture studied.

**Connections:** Connect with other disciplines and acquire information.

Standard 3.1: Students reinforce and further their knowledge of other disciplines through the foreign language.

Standard 3.2: Students acquire information and recognize the distinctive viewpoints that are only available through the foreign language and its cultures.

**Comparisons:** Develop insight into the nature of language and culture.

Standard 4.1: Students demonstrate understanding of the nature of language through comparisons of the language studied and their own.

Standard 4.2: Students demonstrate understanding of the concept of culture through comparisons of the cultures studied and their own.

**Communities:** Participate in multilingual communities at home and around the world.

Standard 5.1: Students demonstrate understanding of the nature of language through comparisons of the language studied and their own.
Standard 5.2: Students demonstrate understanding of the concept of culture through comparisons of the cultures studied and their own (p. 9).

Although the standards are learner-centered and describe what learners should be able to do and accomplish inside and outside of the classroom, they have direct implications regarding teaching and how best to achieve the goals set out by the standards. For example, ‘Sample Progress indicators’ for grades 4, 8, and 12 are included with the standards to aid teachers, and students, in assessing linguistic and cultural development. Moreover, the standards state that “professional development for practicing teachers will also be crucial, and the message of standards must permeate those learning experiences as well” (p. 15).

ACTFL has also produced a set of standards specifically for foreign language teachers in conjunction with the National Council on the Accreditation of Teacher Education (NCATE). These six standards are similar to the ACTFL standards for learners and reflect the knowledge and skills an effective teacher should aspire to obtain. The full-length version of the standards includes each content standard, two to three supporting standards, and supporting explanations and rubrics for each supporting standard. These standards specifically reflect the skills set and knowledge base that foreign language teachers should aim for—according to ACTFL and NCATE. An abbreviated version of the standards is presented below along with supporting standards:

Standard 1: Language, Linguistics, Comparisons

Standard 1.a. Demonstrating Language Proficiency.

Standard 1.c. Identifying Language Comparisons.

Standard 2: Cultures, Literatures, Cross-Disciplinary Concepts

Standard 2.a. Demonstrating Cultural Understandings.

Standard 2.b. Demonstrating Understanding of Literary and Cultural Texts and Traditions.

Standard 2.c. Integrating Other Disciplines in Instruction.

Standard 3: Language Acquisition Theories and Instructional Practices

Standard 3.a. Understanding Language Acquisition and Creating a Supportive Classroom.


Standard 4: Integration Of Standards into Curriculum and Instruction

Standard 4.a. Understanding and Integrating Standards in Planning.

Standard 4.b. Integrating Standards in Instruction.


Standard 5: Assessment of Languages and Cultures

Standard 5.a. Knowing Assessment Models and Using them Appropriately.

Standard 5.b. Reflecting on Assessment.

Standard 5.c. Reporting Assessment Results.

Standard 6: Professionalism

Standard 6.b. Knowing the Value of Foreign Language Learning (NCATE Website).

The international organization of Teachers of English to Speakers of Other Languages (TESOL) has drafted several sets of standards to aid teachers and learners in formulating an idea of what to aim for in their language learning pursuits. Three sets of standards that have been produced by TESOL include “ESL Standards for Pre-K-12 Students,” “TESOL/NCATE Standards for P-12 Teacher Education Programs,” and “Standards for Teachers of Adult Learners.” The connection between effective teaching and standards is made more explicit by TESOL than by ACTFL. In prefacing the standards to the pre-K-12 standards the website states that “TESOL’s standards for pre-K-12 ESL students represent a starting point for developing effective [emphasis mine] and equitable education for ESOL students” (TESOL Website—ESL Standards for Pre-K-12 Students).

In the introduction to the teacher education standards the website refers to the notion of “consistency” in preparing ESL teachers across the country which assumes that not all approaches are equally effective. It claims that “the TESOL/NCATE Standards for P-12 Teacher Education Programs standards address the need for consistency [emphasis mine] throughout the United States in how teachers are prepared to teach English as a second language (ESL) to children in P-12 schools” (TESOL Website—TESOL/NCATE Standards for P-12 Teacher Education Programs).

The standards for teachers of adult learners incorporate eight general domains, each with an accompanying standard and performance indicators. The domains include
planning, instructing, assessing, identity and setting, language, learning, content, and commitment to professionalism. Each domain has a numbered standard, an explanatory narrative, and performance indicators for teachers to use as a guide. The standards do not prescribe specific exercises or concrete, low-inference behaviors for the classroom but they do provide a lengthy description of what effective ESL teachers do and what attributes they possess. One performance indicator under Standard 1 in the planning domain states that teachers should “integrate learner needs in planning” (TESOL Website—Standards for Teachers of Adult Learners).

This sort of mandate may be interpreted very differently by each individual teacher according to the environment and make-up of the class. The adult learner standards as cited above have since been withdrawn from TESOL’s website and are no longer available as they have commissioned a review and re-drafting of the standards. The fact that TESOL and other such organizations periodically organize entire committees to update and revise their standards bespeaks the dynamic nature of foreign language teaching.

As the percentage of second language teachers who join professional organizations increases and as they become indoctrinated in the standards that these organizations espouse, the influence of such standards on teachers’ notions of effective teaching becomes stronger. As mentioned previously, the published standards refer to the impact they are having on curricular decisions and assessment practices. Furthermore, educational institutions who seek accreditation for their teacher training programs from organizations like NCATE are more likely to hold up official standards as the measuring
stick against which curricular and classroom practices will be compared. Thus, the impact of the standards movement on what the field at large considers effective foreign language teaching is unmistakable and far-reaching.

STUDENTS’ AND TEACHERS’ PERCEPTIONS OF L2 TEACHING AND LEARNING

In the early 1980’s, Horwitz (1981)—a pioneer in the field of foreign language students’ belief systems—developed a questionnaire entitled Beliefs About Language Learning Inventory (BALLI) meant to examine students’ beliefs concerning second language learning. After having administered the BALLI and the Foreign Language Attitude Survey (FLAS) (De Garcia, Reynolds, & Savignon, 1976) on several occasions with pre-service language teachers in her methods class, Horwitz (1985) reports on representative findings on her students’ attitudes and beliefs regarding foreign language learning. Several interesting results came out of her research with pre-service teachers and their belief systems: 1) they felt that it is the teacher’s responsibility to increase student motivation regarding language learning, 2) they expressed strong feelings regarding how teachers should deal with student responses, 3) although they agreed with the idea of communicative competence they disagreed on how to achieve it, 4) the students mostly agreed with the concept of foreign language aptitude, believing that some people have a special ability to learn languages—especially children, 5) the methods students expressed their belief that some languages are easier to learn than others, 6) they felt that language learning was a unique academic subject and a special kind of learning
different from other disciplines, and, 7) many students felt that it was key to listen and repeat the language frequently and to practice in the language laboratory.

In a later study, Horwitz (1988) surveyed 241 first semester foreign language students from German (80 students), French (63 students), and Spanish (98 students) classes. The version of the BALLI that was used contained 34 items distributed across five areas with a 5-point Likert scale ranging from strongly disagree to strongly agree. These areas included the following: 1) difficulty of language learning, 2) foreign language aptitude, 3) the nature of language learning, 4) learning and communication strategies, and 5) motivations and expectations. In relation to the difficulty of language learning Horwitz found that students believed there was indeed a difficulty hierarchy among languages with some being harder to learn than others. Another noteworthy finding that surfaced from the data was the feeling among students that a maximum of two years was enough to learn another language. Students also felt rather confident about their ability to ultimately be able to speak the language fluently. Horwitz responds to these beliefs by encouraging teachers to discuss with their students realistic timelines for acquiring another language and the importance not to solely focus on the upper levels of proficiency in language learning.

Similar to the pre-service teachers, first semester students felt that some individuals have special abilities in learning another language while others do not. Students concurred with the commonly held belief that children are more capable at learning languages than adults. However, students’ responses did not coincide with two other common beliefs: 1) that those who are good at math or science are not good
language learners, and, 2) that women are better than men at language learning. Students were not convinced of Americans’ abilities to learn languages as only one student from the entire sample agreed with the statement “Americans are good at learning foreign languages.”

Regarding the nature of language learning, students tended to agree that language learning was different from other school subjects but presented a segmented, narrowly-defined perspective on the key to language learning. More specifically, many students voiced the opinion that learning a language is primarily based on vocabulary and grammar learning. Moreover, learners from the German and Spanish groups felt that translation was crucial in facilitating language learning while the French students did not. Horwitz (1988) points out that these views may lead to students investing too much time memorizing isolated, decontextualized vocabulary items and grammar rules while using the first language as a mediator. Of particular interest is the fact that students from all three language groups demonstrated the same response patterns to nearly all items on the questionnaire with the only exception being the French group’s perspective on translation. That is to say that the French learners did not feel translation was essential for language learning. Two possible explanations for this are put forth by Horwitz, both grounded in classroom instruction: 1) the type of classroom instruction they were exposed to went against and changed students’ previously held beliefs on the role of translation in language learning, and, 2) students were able to infer from instructional practices that translation was not a valuable approach to effective language learning.
In regards to learning and communication strategies, students felt strongly about the role of repetition in language learning and the use of the language laboratory; however, the results were mixed concerning communicative language teaching. Although students agreed that one should feel free to guess the meaning of a word in the target language, they also expressed concern over fossilization. They felt that uncorrected errors made early on at the beginning level would be difficult to eradicate later. Horwitz (1988) argues that those students who value highly grammar-focused language learning might struggle to adjust to the increasingly prevalent communicative approaches to second language pedagogy that emphasize comprehension of meaning over accuracy. At this point Horwitz makes reference to the clash between students’ and teachers’ beliefs and the ramifications of such differences as mentioned in Chapter 1.

The last area that students responded to related to motivations and expectations of language learning. Only a minority of participants correlated language learning with increased employment opportunities while a much larger number felt that they would at least have many chances to use the language. Not surprisingly, the Spanish students felt most optimistic about their future opportunities with the language as compared to the German and French students. Most students felt that Americans in general did not consider language learning important. Finally, Horwitz (1988) found that both instrumental, or utilitarian, and integrative motivation among students were demonstrated to be only moderate. Evidently, many of the students did not believe learning an additional language would lead to a better job, nor did they express the desire to become familiar with individuals from other cultures. This last finding led Horwitz to make an
insightful observation that “It is not surprising then that when language study proves to
be more difficult and time consuming than they originally estimated, the majority will
probably quit language study as soon as possible” (p. 291).

In discussing the value of the study, Horwitz (1988) comments on the pedagogical
implications for foreign language classrooms. The study was designed to simply identify
students’ belief systems but she also emphasizes the fact that certain commonly held
beliefs, at least within this sample of students, are quite disconcerting and need to be
addressed. Foremost among those are 1) the feeling among almost 40% that one could
become fluent in another language in two years or less, and, 2) the belief among over
60% of Spanish and German students that learning is mostly a matter of translating from
English. The author recommends that teachers need, first, to make students aware of their
own notions of language learning along with the possible consequences and, second, to
include discussions of language learning as part of classroom instruction—a proposal to
be taken up in more depth in Chapter 5.

Kern’s (1995a) administration of Horwitz’s BALLI questionnaire to students and
teachers represents one of the few studies that directly compares students’ beliefs about
language learning to their teachers’ with an identical instrument. Kern raises a question
very similar to the objectives driving the current research, namely, “To what degree do
foreign language students’ beliefs about language learning correspond to those of their
teachers?” (p. 72). The study included 288 first and second semester French students and
their 12 instructors and was administered twice over the course of the semester—once on
the second day of classes and again during the last week of the semester. Any students
who missed either the first or second administration of the survey were not included in
the results and data analysis. The version administered to teachers was slightly altered to
reflect teachers’ perspectives, i.e., “the language I teach” instead of “the language I am
learning,” and contained fewer items as some were excluded after being deemed
irrelevant to teachers, i.e., “If I get to speak this language well, I will have many
opportunities to use it.”

Kern’s (1995a) findings on students’ responses were similar to the results from
Horwitz’s study with only slight variation. Kern found that students in his study also felt
that language learning was a relatively speedy process and that learning a language was
quite different from other academic subjects. Additionally, Kern’s participants did not
perceive that Americans felt language learning to be important and less than half of the
students thought a knowledge of the foreign culture was essential. Students’ opinions
regarding strategies of communication and learning reflected both traditional approaches
to language pedagogy, e.g., mechanical practice and repetition, as well as more current
strategies such as the communicative approach with its tolerance for errors in oral and
written production tasks.

In regard to the comparison between teachers’ and students’ responses, Kern
reports a very high overall correlation between all students and all teachers of .93 across
the 27 overlapping items. However, when the data is analyzed by student with each
student’s responses being compared to his/her own teacher, correlations ranged from .00
to .80. The overall comparison demonstrated differences in perspectives on pronunciation
or accent, learning grammar rules, difficulty of speaking versus listening, and translation.
The greatest differences between students and their individual instructors resulted on the items where the greatest differences were found between teachers. These items related to the importance of cultural knowledge in language learning, whether learning the language in the target culture was more effective, the amount of time needed to become fluent in a foreign language, the need to correct errors to avoid fossilization, and the relative ease of reading and writing compared to listening and speaking. In summary, Kern points out the following:

when viewed from a global perspective, the instructors and students appear to have relatively few conflicts with their beliefs. But when the analysis shifts to the level of the individual classroom, one finds frequent instances of substantial difference in opinion (p. 78).

Another question this study attempts to answer is whether students’ beliefs tended to change over the course of the semester, possibly suggesting teacher and/or classroom influence? When analyzed overall, students’ beliefs appeared quite stable from the beginning to the end of the semester. On the contrary, when individual analyses are conducted on students’ responses, the results indicate that on any given item 35 to 59 percent of students changed their responses over the 15-week period. If all items are considered, 52 percent of individual responses changed over the course of the semester.

Among the many shifts that took place in students’ answers, one of the most noteworthy was in response to an item about fossilization and error correction. In reaction to the statement, “If you are allowed to make mistakes in the beginning it will be hard to get rid of them later on,” 37 percent of students moved toward greater agreement with 11
percent moving 2 or more points on the 5-point scale. Hence, in this particular case, students’ responses were in less agreement with their teachers at the end of the semester than at the beginning. Kern (1995a) hypothesizes that these students became more aware of their errors and the difficulty of overcoming them over the course of the semester. This teacher-student mismatch in beliefs seems to indicate that teachers’ beliefs regarding error correction bear less of an impact on students’ beliefs than concrete practices. Other areas that shifted to more agreement were issues relating to the importance of accent, the ease of speaking as compared to listening, and the pivotal role of grammar learning. This final shift in student responses contrasted greatly with teachers who felt that learning grammar rules was not crucial to language learning. The fact that differences between students’ and teachers’ beliefs existed both early in the semester and later on, and in some cases even increased, seems to provide further evidence that beliefs and attitudes regarding language learning are difficult to change.

In discussing implications of his study, Kern (1995a) refers to the need of open communication between students and teachers regarding language learning and teaching beliefs and practices. He specifically points to the responsibility teachers have to encourage students while providing them with realistic expectations for language learning. Kern, like Horwitz (1988), suggests that teachers share what they believe with their students and the rationale driving their teaching. He argues that “by listening closely to our students, by identifying mismatches in beliefs, and by clearly explaining why we do what we do in the classroom it may be possible to significantly allay student frustration” (p. 81).
One of the key differences between the present instrumentation and the BALLI developed by Horwitz (1985) is the heavy focus on concrete practices instead of general principles of language learning. Rather than query students and teachers about abstract principles of language learning, the current study specifies concrete teaching practices considered relevant in contemporary foreign language courses. A recent study conducted by Reber (2001) includes an extensive profile of teachers’ beliefs concerning both abstract principles of foreign language pedagogy and second language acquisition theory.

Reber’s (2001) dissertation research with post-secondary FL teachers throughout the western United States represents one of the most salient and thorough studies into foreign language teachers’ perceptions of abstract SLA theory as well as pedagogical practice. An 80-item questionnaire was mailed to 1,000 post-secondary instructors, 950 of whom were members of ACTFL. After an extensive literature review to determine salient issues in second and foreign language learning, Reber personally developed an 80-item instrument distributed over nine different categories: 1) ACTFL’s Standards for Foreign Language Learning, i.e., Communication, Cultures, Connections, Comparisons, Communities; 2) corrective feedback; 3) theories and teacher behaviors related to communicative approaches; 4) focus on form in classroom SLA; 5) individual learner differences in FL learning; 6) strategies for foreign language learning; 7) theories about SLA; 8) teacher qualifications; 9) assessment in foreign language teaching. These categories gave shape to each individual item included in the questionnaire, e.g., Category: Corrective Feedback; Item: The effective foreign language teacher corrects errors as soon as possible after they occur.
The 457 foreign language teachers who responded to the mailed questionnaire agreed with the majority of items related to the ACTFL Standards, communicative language teaching, small group work, and strategies for FL learning. However, only one of the eight items pertaining to the teaching of grammar and two of the eight items relative to assessment in FL teaching reflected a high level of agreement. Furthermore, 14 of the items on the questionnaire, almost 20%, did not receive a majority agreement or disagreement. The author proposes that this response pattern may be indicative of controversial areas in SLA and L2 teaching, such as error correction, Krashen’s Monitor Model, focus on grammatical form in classroom, individual learner differences, and assessment. As evidenced by Reber’s research, certain controversies within SLA theory and second language pedagogy have yet to be solved and continue to generate debate—even among those with formal training in the field.

Brosh (1996) also collected data on the perceived characteristics of an effective language teacher from not only FL teachers but also FL students. Two hundred teachers and 406 ninth-grade students in Tel Aviv, Israel were randomly selected to complete the survey. Unlike Reber’s (2001) instrument which used Likert-type questions, Brosh (1996) provided teachers and students with a list of 20 characteristics and asked them to choose the three most important and to rank order those three. The first three items are included to provide a sense of the instrument: 1.) Prepares and organizes the lesson, 2.) Acquainted with the curriculum, 3.) Helps students after class time (p. 136). Of the 20 total items, only five were specific to language learning.
The results of Brosh’s (1996) questionnaire demonstrated that students’ and teachers’ perceptions were largely homogeneous. The item that both groups chose as most important was the teacher’s command of the target language. The second most crucial factor for both students and teachers was the instructor’s ability to transmit knowledge comprehensibly while motivating students to do their best. Students’ choice for the third most important characteristic differed with teachers’ as they indicated the essential need of being treated equitably and fairly while teachers ranked the ability to provide students with successful experiences as third.

In two separate studies, Schulz (1996; 2001) compared students’ and teachers’ attitudes toward grammar teaching and error correction, both domestically and abroad. In the first study (1996), the researcher distributed a mailed, Likert-type questionnaire to 213 FL instructors at the University of Arizona, 92 of which responded resulting in a 43% response rate, and 824 students to determine each group’s perspective on grammar teaching and error correction. Students and teachers responded to seven questions regarding grammar teaching and five questions regarding error correction by indicating the degree to which they agreed or disagreed on a 5-point scale with statements such as the following: “For adolescents or adults, the formal study of grammar is essential to eventual mastery of a FL/L2 when language learning is limited to the classroom.” The items for students and teachers were not identical as the preceding item was only presented to teachers while students responded to a related item worded in a slightly different manner: “The formal study of grammar is essential to eventual mastery of a foreign language.”
In general, Schulz found that students’ attitudes toward formal grammar study were more favorable than their teachers’ and that students tend to be more accepting of receiving grammar instruction than their instructors are of giving it. Although this finding does not imply that students enjoy grammar learning, it does highlight the strong conviction they have of the central role grammar teaching and learning have in second language learning. Students from less commonly taught languages (LCTLs) actually expressed a higher degree of enjoyment in studying grammar as compared to their counterparts in the commonly taught languages (CTLs). Another interesting difference that arose between language groups was the belief among almost 20% more LCTL than CTL teachers that the “study of grammar helps in learning an FL” (p. 346).

Much like students’ perspectives on grammar teaching, their attitudes concerning error correction were quite positive. In response to the statement “Generally, when students make errors in speaking the TL, they should be corrected” 90 percent of students from both language groups agreed while only 34% of CTL teachers and 50% of LCTL teachers agreed. Schulz (1996) summarizes her findings by stating that students are generally in favor of grammar instruction while many teachers are in disagreement concerning the role of explicit grammar instruction—a finding she calls “perturbing” (p. 348). Three explanations are put forth to account for students’ generally favorable perspectives on grammar teaching and error correction: 1) students’ ideas may be purely based on myths regarding the importance of grammar study passed on from one generation of learners to the next, 2) students may be influenced by the preponderance of grammar-based curricula and discrete-point testing practices still found in FL classrooms,
and 3) students’ opinions may be backed by personal experiences causing them to feel that their foreign language learning is aided by grammar learning and corrective feedback.

Like Kern (1995a) and Horwitz (1988), Schulz (1996) recommends further research into the match or mismatch between students’ and teachers’ belief systems in the FL classroom. She makes an insightful observation that the degree of effectiveness of error correction strategies adopted by teachers depends, partly, on the attitude of the students’ toward those strategies. This may hold true not only for error correction but other common practices performed in the FL classroom. Schulz foreshadows the need for the current study by stating:

. . . it might well be wise to explore the fit of learner and teacher beliefs . . .

While opinions alone do not necessarily reflect the actual cognitive processes that go on in language acquisition, perceptions do influence reality. Indeed, some would argue that perception is reality for the individual learner. Students whose instructional expectations are not met may consciously or subconsciously question the credibility of the teacher and/or the instructional approach . . . Such lack of pedagogical face validity could affect learners’ motivation . . . (p. 349). She adds that teachers might want to explore their students’ beliefs and expectations and compare them with their own.

Schulz (2001) conducted a similar study in Colombia with teachers and students and found comparable results. In the Colombian context, teachers and students felt, even more so than their American counterparts, that grammar instruction plays a vital role in
improving communicative ability and “clearly have greater faith in the benefit of grammar study than U.S. students” (Schulz, 2001, p. 247).

Levine (2003) surveyed 600 foreign language students and 163 foreign language teachers on their opinions regarding their estimations of quantity of target language (TL) use in various classroom contexts in university FL classes, beliefs about the importance of TL use, and beliefs about student anxiety due to TL use. The student questionnaire contained seven sections totalling 63 items and the instructor version contained six sections with a total of 58 items. The findings from the on-line, anonymous questionnaire data revealed some intuitive results and others not so intuitive.

As may be expected, Levine found that both students and teachers felt that students used the L2 less than their teachers and did so even less when interacting with other students as compared to their interactions with the teacher. The TL was perceived to be used more during topic, theme-based communication than communication about grammar and even more still than during interactions about tests and assignments. Only 40% of students reported that using the foreign language made them feel anxious and a large majority, 63%, felt that it was rewarding and a worthwhile challenge to have to use the TL. Possibly the most surprising finding from Levine’s study was that students who perceived higher levels of TL use in their classrooms also reported lower levels of anxiety regarding TL use. Furthermore, the teachers who responded to the questionnaire appeared to perceive higher levels of L2-related anxiety by students than did students themselves.
The research cited here has contributed to our understanding of students’ and teachers’ perceptions of certain aspects of L2 teaching and learning at a general level. Some of the studies focused on general learning theories and not specific, concrete practices (Horwitz, 1988; Horwitz, 1985; Kern, 1995a) while others did focus on particular teaching behaviors but did so in a manner that did not allow comparisons between individual teachers and their students (Levine, 2003; Schulz, 1996; Schulz, 2001). Questions still remain such as “How do Teacher X’s perceptions of concrete behavior Y compare to students’ perceptions of Teacher X?,” and “What degree of similarity or difference between teachers’ and students’ beliefs would emerge across several dozen different teachers and their specific classes?” Furthermore, studies such as those cited above inform the field regarding perceptions at a general level but do not allow students and teachers to concretely evaluate, not just observe or report on, specific teaching practices as performed in their own classes.

THE VALIDITY OF STUDENT RATINGS

As of 1993, nearly 90% of American colleges were using student evaluations (SE) to assess faculty teaching (Seldin, 1993). More recently, Kulik (2001) reports that at his institution student ratings are currently used in hiring new faculty, in reviewing current faculty, in determining faculty promotion and tenure, in qualifying for accreditation, in selecting faculty for awards and honors, and in assigning courses to faculty. In their study of business school administrators, Read, Rama, & Raghunandan (2001) report that SEs receive more than half of the overall weight in evaluating teaching.
Whether it be society’s response to student-centered educational theories, a result of political rhetoric demanding educational accountability, or simply a question of convenience and money, student evaluations continue to play a central role in current evaluative practices of college teaching—regardless of discipline. The assumption underlying the use of student evaluations is that students are capable of making valid assessments of their teachers’ instruction. An even more basic assumption is that students are aware of what effective teaching is in a given discipline. With the increased integration of student evaluations in professional advancement and the weighty consequences for educators, the question remains, “How valid are student evaluations of teaching?”

In spite of the long-standing use of student evaluations in modern educational practices, it is the aforementioned question that causes some faculty to harbor serious misgivings regarding student ratings. Theall and Franklin (2001) astutely point out that for many faculty the idea of allowing someone else to “determine the quality of one’s work is threatening, and when the evaluators of the work are not considered to be as qualified as the evaluatee, anxiety and resistance can increase” (p. 47). Consequently, much research has aimed at discounting the usefulness of SEs concluding that student evaluations are invalid, unreliable, and biased.

During the late 70’s and early 80’s when research on student ratings was flourishing (Eiszler, 2002), one of the foremost authorities on student evaluation research, Aleamoni (1981), identified several common concerns of teachers: 1-inconsistency of student evaluations due to immaturity and lack of experience; 2-only qualified colleagues
are competent to judge a teacher’s effectiveness; 3-most student ratings of professors reflect a popularity contest; 4-students need to be separated from the course and the university for several years to make an accurate evaluation; 5-student rating forms are both unreliable and invalid; 6-other variables often cited: size of the class, sex of the student and instructor, time of day the course was offered, etc.; 7-students’ grades are highly correlated with their ratings of the course and the instructor; 8- the unlikelihood that student evaluations can be used to improve instruction. In a recent version of the article, Aleamoni (1999) included additional concerns of many faculty: whether the course is an elective or required; whether the course is part of students’ major or minor; whether students are underclassmen or upperclassmen; whether the instructor is assistant, associate, or full professor; whether students’ grades reflect their ratings; whether certain disciplines receive higher or lower ratings than others; whether a single, general item accurately reflects instructional effectiveness; whether student evaluations can be used to improve instruction. Aleamoni systematically refutes many concerns by citing more current research from the field and recommends methods of dealing with valid concerns while arguing that none of the empirical research invalidates student ratings.

In spite of Aleamoni’s (1999) assertions regarding student ratings, faculty reservations still persist and have been addressed by Kulik (2001), Theall & Franklin (2001), and Ory & Ryan (2001), each in a slightly different way. Kulik argues that the validity of student ratings depends on how closely they correlate with teaching effectiveness scores. Thus, he systematically demonstrates how research has demonstrated a positive correlation between SEs and four credible indicators of teaching
effectiveness (student learning, student comments, alumni ratings, and ratings from outside observers). Furthermore, Kulik systematically analyzes and refutes the findings from five frequently cited studies that appear to prove the invalid nature of SEs (Rodin & Rodin, 1972; Naftulin, Ware, & Donnelly, 1973; Ambady & Rosenthal, 1992; Greenwald & Gillmore, 1997; Williams & Ceci, 1997).

Possibly the strongest and most commonly cited critique of student ratings was made by Rodin & Rodin (1972) who found a -.75 (negative) correlation between student ratings and student learning in an undergraduate calculus course. Their findings were published in the prestigious journal *Science* and led many to believe that student evaluations were completely invalid. Kulik (2001) notes that the evaluations were of the 11 teaching assistants, who did only some teaching, and not the actual instructor. In addition, the method for assessing learning was comprised of students’ performance on 40 exam questions given separately throughout the semester after each unit. Each teaching assistant graded their own students’ problems; hence, on the SEs the effect of different grading standards was confounded with the assistants’ teaching performance.

The scope of this section has only allowed for an analysis of the Rodin and Rodin study, but as Kulik argues each study of those cited above (Naftulin, Ware, & Donnelly, 1973; Ambady & Rosenthal, 1992; Greenwald & Gillmore, 1997; Williams & Ceci, 1997) had serious methodological problems affecting the power of statistical results and the generalizability of the findings. Moreover, several of the studies produced anomalous results, including Rodin & Rodin’s study, that contradict a multitude of other studies (Kulik, 2001).
Ory (2001) presents his overview of SE research by way of a conversation with colleagues who express doubts concerning the accuracy of SEs. Ory addresses his colleagues concerns on several issues that may affect the validity of SEs, ranging from elective vs. required courses, class size, discipline area, student/teacher gender, and expected grades. Some of his most noteworthy conclusions are the following: ratings in elective courses tend to be higher than required courses, higher-level courses lead to higher ratings although only marginally so, students expecting high grades give higher ratings, the gender of the instructor demonstrates no significant relationship to SEs although students tend to give higher ratings to teachers of the same sex, class size is not a significant source of bias, and certain disciplines such as the arts and humanities received higher ratings than hard sciences such as the mathematical and physical sciences. In the areas where research has shown differential ratings, Ory recommends the creation of separate norms.

Theall & Franklin (2001) synthesize the literature by answering specific questions addressing myths and misconceptions regarding student evaluations, several of which will be treated briefly here. First, are students qualified to evaluate their teachers and their instruction? While the myth says no, the authors claim that in general they are. One area that the authors deem students unqualified to rate, however, is the teacher’s depth of knowledge. Ory & Ryan (2001) include course content as another area students cannot adequately assess. Second, are ratings based exclusively on “popularity?” The authors respond by stating that the myth here is that popular teachers are poor teachers, which has no backing in the research. Third, are ratings related to learning? Theall & Franklin
conclude that those who learned more gave higher ratings. Fourth, are student ratings reliable? Student ratings appear to be remarkably consistent within classes, across classes, and over time. Fifth, does gender make a difference? When gender is isolated from other factors such as class size, discipline, and level, there do not seem to be strong patterns of gender bias. Sixth, are ratings affected by situational variables? The researchers propose that SEs are not strongly affected by situation although they admit that SEs from a large, out-of-major course in the physical sciences may be lower than those in an upper-level elective, in-major course. Finally, do students rate teachers on the basis of expected or given grades? The argument given is that good teaching leads to learning which leads to high grades and satisfaction. Thus, it follows that good students expecting good grades would give high ratings.

Theall & Franklin (2001) point out that this last question regarding the correlation between expected grades and student ratings has been the source of much contentious discussion. Chambers & Schmitt (2002) considered this issue in depth with a study testing the validity of the leniency hypothesis. This hypothesis simply suggests that students who expect high grades from their teachers will in turn rate them higher. Chambers & Schmitt further hypothesize that students will make comparisons between their exerted effort and their expected evaluation in one instance, or class, with their efforts and evaluation in another and then make corresponding evaluations as to the equity of their received assessment, i.e. equity theory. In their large-scale study of 224 classes over five years, the researchers compared students’ grades with their ratings of a specific class and also compared those grades and ratings to other classes. Indeed,
students who expected higher grades also rated their instructors in a similar fashion. Eiszler (2002) found similar results in investigating grade inflation as he found that faculty received higher ratings during the same semesters in which they gave a higher percentage of A and A- grades. However, the statistical analyses from Chambers & Schmitt’s study demonstrated that the leniency hypothesis alone could not explain a significant amount of the variance leading the authors to conclude that other factors must be at work.

Research studies such as those mentioned above, whose aim is to determine the validity of student ratings, can be divided into five categories: multisection, multitrait-multimethod, bias, laboratory, and dimensionality (Ory & Ryan, 2001). Multisection studies look at the ratings given to several instructors teaching different sections of the same course. Multitrait-multimethod research determines the fit between student ratings and other criteria of effective teaching, such as peer-ratings, and self-evaluation. Studies on bias correlate with any number of variables extraneous to the rating process, i.e., student, instructor, or course characteristics. Laboratory studies are typified by experimenter-controlled variables that take place outside of a natural, classroom-based context. The final approach to studying student ratings attempts to identify a common set of factors along a single dimension within the research at large, as in a meta-analysis of multiple studies.

An important caveat must be made as to research design and methodology in SE research when considering the conclusions of different scholars in the literature, especially those that argue in favor of student ratings. Many times they claim that the
effect of a given variable, i.e., gender or class size, may only be marginal. However, in many cases the studies that the authors reviewed included large numbers of students who completed questionnaires comprised of Likert-type items on a 4 or 5-point scale. The tendency of large numbers to regress toward the mean and neutralize significant numbers of outliers may lead to misinterpreted data. If a first-year, female professor who has two or three male students in a class of 12 rate her extremely low as compared to the remaining 9 or 10 students, her overall class evaluation will not be truly representative of her teaching effectiveness. In this case the instructor has little or no previous ratings with which to compare. Student evaluation research, as so many other fields of quantitative inquiry in the social sciences, often times works with averaged scores leaving individual trends unattended.

In considering a new framework for measuring the validity of student evaluations, Ory & Ryan (2001) emphasize the importance of consequential validity, or the way in which student ratings are interpreted and used. Beyond the detrimental effect personal preferences and biases may have on student evaluations, the way in which student evaluations are implemented into faculty evaluation and assessment remains of utmost importance. Student ratings for the most part are valid in spite of certain tendencies and trends identified above. However, many faculty continue to challenge their utility in assessing teaching. One plausible explanation for why many have vehemently decried the validity of student ratings could be that they feel SEs are given too much weight in the summative evaluation of teaching. For this very reason, those scholars who defend the validity of SEs argue repeatedly that SEs should represent only one facet of a multi-
dimensional approach to evaluating teaching. They further argue that SEs must not only serve to inform summative, high-stakes decisions, but they also must serve to instigate concrete changes of a formative nature (Arreola, 2000, Ory & Ryan, 2001, Theall & Franklin, 2001).

Up to this point, the discussion surrounding the validity of SEs has not distinguished between the evaluations of language students and students of other disciplines. Although many of the threats to validity mentioned above also apply to language students since they do not represent a significantly different population demographically, surely the unique nature of the foreign or second language classroom presupposes other important considerations, some of which have already been mentioned.

STUDENT EVALUATIONS VS. TEACHER SELF-EVALUATIONS

Unfortunately, to this author’s knowledge, except for a small pilot study (Beaudrie, Brown, Thompson, 2004) which will be commented on later, no research study has correlated the evaluations of specific classes of foreign and second language students to their own teachers’ self-evaluations via a similar evaluation instrument. That is to say that students’ and teachers’ evaluations of L2 teaching using the exact same object of observation and a similar assessment instrument remain largely unattended. Studies outside the field of language teaching have produced such comparisons.

Moore (1996) takes a step in that direction by administering the exact same evaluation instrument to both teachers, in this case graduate student instructors, and their students. As part of his dissertation research, Moore (1996) investigated the correlation between teachers’ and students’ perceptions of teaching effectiveness. Moore’s study
included 129 graduate teaching assistants and their 3,088 students from various
departments across a large university in the southeastern United States. The results
indicated that students were consistent in evaluating their graduate student instructors
lower than the instructors rated themselves on the same instrument. A positive correlation
was found between prior teaching preparation and experience and the perceptions of
students and teachers. Although this study was not conducted exclusively in the language
classroom, it demonstrates the disconnect that may occur between teachers and students
regarding an assessment of teaching.

Self-assessment is used widely in professional development for teachers of all
levels and disciplines, including foreign and second language learning. Bailey, Curtis, &
Nunan (2001) dedicate an entire textbook for language teachers to orient them on how to
develop themselves professionally using self-reflection and self-assessment. However,
not much work has been done to correlate student ratings with teachers’ self-ratings,
especially when using a similar instrument. Blackburn & Clark (1975) conducted a study
at a small college that compared self-assessment, student assessment, and peer
assessment. They found that instructors’ self-ratings of teaching effectiveness correlated
only slightly (r = .19) with their students. A closer look at the research design indicates
that teachers were asked to assess their teaching effectiveness generally—not in reference
to a specific class—while students responded to the standard, end-of-semester evaluation.
Marsh, Overall, & Kesler (1979) improved upon this design by administering the exact
same instrument to instructors that their students filled out, except that it was worded in
the first person. They found significant correlations (median $r = .49$) between student evaluations of teachers and teachers’ self-assessments.

Feldman (1989) provides a meta-analysis of numerous studies into instructional effectiveness where each included comparisons of at least one of the following groups of raters: current students, former students, self-ratings, colleague ratings, administrator’s ratings, and external observer ratings. He separates his analysis into two distinct types of similarity: *relative* and *absolute*. Relative similarity refers to the correlation coefficient, or product-moment correlation ($r$), between different groups’ ratings of instruction so that one group may correlate perfectly with another ($r = +1.00$) even though their evaluations were an entire step apart on the rating scale. Absolute similarity, then, explores the extent to which each group’s raw rating numbers are exactly alike, e.g., $t$-test.

In his meta-analysis, Feldman discovered that the 19 studies that made up the ‘current students’ vs. ‘self’ group resulted in an average correlation of $r = +.29$—a rank of six out of eight among comparison groups where meaningful averages could be calculated. This same comparison group, ‘current students’ vs. ‘self’, resulted in the largest range of correlations among all compared groups. One study found a correlation of $r = .75$ between students’ ratings and teachers’ self-ratings while another study reported a correlation of $r = -.17$ (negative).

In comparing the *absolute* similarity of teachers’ self-ratings with their current students, Feldman’s meta-analysis of 10 studies did not produce a significant difference between the two groups. Feldman warns that this should not be taken to mean that individual teachers rated themselves the same as their students. Since the statistic
reported is an average, Feldman highlights the fact that some teachers rated themselves higher than their students while others rated themselves lower and some ratings were similar between teacher and students. Neither the Feldman (1989) meta-analysis nor the other two studies mentioned above (Marsh, Overall, & Kesler, 1979; Blackburn & Clark, 1975) were conducted solely in the second or foreign language classroom.

A small-scale, pilot study conducted by Beaudrie, Brown, and Thompson (2004) triangulated self-assessment, peer assessment, and student evaluations in the Spanish foreign language classroom in an attempt to determine the degree of similarity or difference between each group’s ratings. Three peer observers attended a total of seven class sessions across the instructor’s two Spanish classes of the same level. After a three-and-a-half week period of time the peer evaluators, the instructor, and both sections of students filled out the exact same evaluation instrument. The evaluation employed a 35-item, Likert-type instrument, specifically designed for use in the language classroom. Statistical analyses demonstrated a significant difference between each group’s evaluation of the teaching. Unlike Moore (1996) who found graduate student instructors overrated themselves, this study found that the graduate student teacher’s self-assessment provided the lowest rating of the three groups. Furthermore, a comparison between the responses on the 35-item instrument and the standard university evaluation form demonstrated significantly different responses raising the question of subject-specific instrumentation in the language classrooms. In a similar fashion, Sojka, Gupt, and Deeter-Schmelz (2002) found that both students and faculty would be interested in “more specific information on evaluations and both seem to prefer questions specifically
relating to the teacher’s area of expertise” (p. 47). Simply put, the similarity or difference specifically between L2 teachers’ self-ratings and their students’ ratings have not been explored extensively enough to allow for any reliable conclusions.

CONCLUSION

This chapter has outlined the development of current notions of effective foreign language teaching, reviewed empirical research into those notions, reported on students’ and teachers’ perceptions of these issues, and, finally, considered the validity of students’ and teachers’ self-evaluations. Chapter 2, then, has provided the necessary background for Chapter 3 which details the current study, its context, its structure, and methods of analysis.
CHAPTER 3
METHODOLOGY

INTRODUCTION

The present study can best be described as exploratory in nature as the intent is to discover the perceptions of two distinct, albeit closely related, groups of stakeholders in the process of foreign language teaching—foreign language teachers and their students. Within the specific geographical and demographic context of this study, the principal objectives are 1) to identify post-secondary L2 students’ and teachers’ perceptions of effective teaching and the frequency of occurrence of certain behaviors in the classroom; 2) to compare these perceptions, and, 3) to compare students’ evaluations of teaching to their instructors’ self-evaluations. Secondarily, the results of this research will allow for a comparison of selected items from the university’s TCE form with the discipline-specific questionnaires used in this study. These objectives led to the research questions as formulated in Chapter 1, which are reproduced here:

Research Question 1

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do students consider to be reflective of effective foreign language teachers?

Research Question 2

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do these students’ teachers consider to be reflective of effective foreign language teachers?
Research Question 3

Based on a similar 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), how do students’ and teachers’ perceptions of effective foreign language teachers coincide or differ?

Research Question 4

How do individual teachers’ perceptions of effective foreign language teachers coincide with or differ from those of their students’?

Research Question 5

How do students’ and teachers’ perceptions of the frequency of occurrence of certain teaching behaviors in their classrooms coincide or differ, both overall and in a teacher-by-teacher comparison on a 21-item, Likert-scale questionnaire (Evaluation Questionnaire)?

Research Question 6

How do teachers’ self-evaluations coincide with or differ from their students’ evaluations of them overall and individually?

Research Question 7

What impact do language type and class level exert on students’ responses to both the 24-item (Effective Teacher) and 21-item (Evaluation) questionnaires?

Research Question 8

How do students’ evaluations of their foreign language teachers on the 21-item language-specific instrument (Evaluation Questionnaire) coincide with or differ from
their evaluation of their teachers on selected items from the general, university-wide
evaluation form, i.e., the Teacher-Course Evaluation (TCE)?

Research Question 9

What relationship exists between students’ response patterns on both the Effective
Teacher and Evaluation Questionnaires and students’ responses to selected items from
the TCE instrument?

The following section will provide a brief summary of the research design and
methodology adopted in this investigation, first by describing the teachers and students
who make up the target population, and, second by detailing the construction of the
instrumentation, its rationale, and its implementation. Finally, the methods of data
collection and the procedures for data analysis will be presented.

PARTICIPANTS

The participants for this study were 49 teachers (12 males, 37 females) mostly
between the ages of 21 and 35 and their students recruited on a volunteer basis from 83
intact 1st- and 2nd-year foreign language classes from nine different languages at the
University of Arizona. These languages included both commonly taught languages—
Spanish, French, German, Italian—and less commonly taught languages—Italian,
Turkish, Japanese, Greek, Arabic, and Hebrew. Only 1st- and 2nd-year classes were
recruited given the nature of the instrumentation and the irrelevance of many items for
upper-division language courses. Teachers were recruited by way of their basic language
program directors or individually with the department’s authorization. Students were
included as intact classes attached to a specific teacher in order to allow for detailed
comparisons of beliefs, perceptions, and evaluations. As the major research objective is to compare students’ and teachers’ perceptions of effective teaching and their concrete evaluations of it, a special effort was made to recruit L2 teachers who taught more than one section of a language to increase student-teacher comparisons, although no teacher was excluded for only teaching one class or for low enrollment.

**Teacher Population.** The foreign language instructors who participated in this study were those who specifically taught 1st- and 2nd-year foreign language classes, regardless of academic rank or experience. Teachers were asked to provide demographic information as part of the teacher information questionnaire. An explanation of the teacher information questionnaire, its five categories and accompanying questions and the rationale behind the items appears in the instrumentation section (see Appendix D for the Teacher Information Questionnaire and Appendix E for summary of teacher demographics). An overwhelming majority of the instructors (78%) were graduate assistants or associates in teaching (GATs) as most language departments employ M.A. and Ph.D. students, not all of whom are graduate students from the same department, to teach lower-level courses. In some cases full and part-time lecturers, adjuncts, tenured, and tenure-track faculty serve as instructors of lower-level classes. Those departments with large basic language programs such as Spanish, French, and German represent a large majority of participating teachers and students although no languages were excluded due to low numbers. Just over half (51%) self-identified as native speakers of the language they taught. As mentioned previously, many of the participating teachers
taught two sections of the same class and some teachers taught three or more sections across two levels.

**Student Population.** A crucial element of the current research design is the ability to make comparisons between students’ and teachers’ perceptions and evaluations, hence, only those students whose teachers participated in the study were included and no students were recruited separately from their teachers. A majority of the students were female (55%) with a large proportion between the ages of 18-23 (90%) with two-thirds being freshman or sophomore.

Many students in basic language courses enroll in such courses to fulfill a general education requirement (58%) imposed by their major. In general, these students are underclassmen who desire to complete the L2 requirement with limited or no plans for further language study. Nevertheless, many of these undergraduate students plan on pursuing a minor or major in the language of study or enrolled purely out of personal interest. Other students who participated were graduate students and non-degree seeking students who were affiliated with the university as employees taking the courses for personal reasons. Similar to the teachers, students were asked to provide demographic information on the student information questionnaire (see Appendix B for reproduction of questionnaire and Appendix C for summary of students’ responses) which is outlined in the following section on instrumentation. The total number of students enrolled in the 83 classes at the end of the semester as reported by each department was 1606. (See Table 3 for information on languages, levels, number of classes, and participating students per class) The number of students who completed each questionnaire varied as
they were administered on two separate class days resulting in different totals due to student attendance and volunteer participation (see Appendix B for Student Information Questionnaire and Appendix C for summary of student demographics).

Table 1. Summary of Participating 1st and 2nd-Year Foreign Language Classes by Language and Level.

<table>
<thead>
<tr>
<th>Language/Level</th>
<th>Number of Classes</th>
<th>Number of Students(^1)</th>
<th>Language/Level</th>
<th>Number of Classes</th>
<th>Number of Students(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>3</td>
<td>63</td>
<td>101</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>102</td>
<td>3</td>
<td>61</td>
<td>102</td>
<td>14</td>
<td>245</td>
</tr>
<tr>
<td>201</td>
<td>2</td>
<td>33</td>
<td>201</td>
<td>12</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>202</td>
<td>8</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>203</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>205</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>206</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>157</td>
<td>Total</td>
<td>39</td>
<td>778</td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
<td>Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>3</td>
<td>58</td>
<td>102</td>
<td>3</td>
<td>47</td>
</tr>
<tr>
<td>102</td>
<td>2</td>
<td>31</td>
<td>402</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>202</td>
<td>2</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>134</td>
<td>Total</td>
<td>4</td>
<td>65</td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td>Turkish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>3</td>
<td>71</td>
<td>102</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>102</td>
<td>5</td>
<td>134</td>
<td>202</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>201</td>
<td>3</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>202</td>
<td>3</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>342</td>
<td>Total</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Italian</td>
<td></td>
<td></td>
<td>Hebrew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>2</td>
<td>36</td>
<td>103B</td>
<td>5</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>203B</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>36</td>
<td>Total</td>
<td>6</td>
<td>77</td>
</tr>
<tr>
<td>Greek</td>
<td>104</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) in some cases numbers may not be exact due to students’ errors in providing matching codes on questionnaires.
GENERAL CONSIDERATIONS OF QUESTIONNAIRE DESIGN

Brown (2001) defines questionnaires as written instruments that contain statements or questions which participants must respond to by writing answers or choosing from options provided by the questionnaire. He differentiates between open- and closed-response formats, the former referring to surveys that allow participants to respond orally or in writing, the latter referring to those that require them to select from available options. Brown outlines many advantages to the closed-response format. First, the closed-response questionnaire generates uniform data in regards to type and degree of specificity. Second, surveys of the closed type are easy to answer and participants are unlikely to skip questions. Third, the ability to express closed-response responses numerically make the data easier to code, analyze, and interpret. This characteristic of the data makes them appear more objective in nature. Finally, the numerical analysis of closed-response data facilitates the testing of reliability and validity measures.

In describing the advantages of questionnaires, Dornyei (2003) points to the extreme efficiency of questionnaires relative to researcher time, researcher effort, and financial resources. Another advantage mentioned by Dornyei is the versatility of questionnaires arguing that they can be used in a diverse set of situations with a variety of people and topics.

Nevertheless, both authors detail the disadvantages of questionnaires that constrain their use. Brown (2001) begins his discussion of disadvantages by referring to one of the main issues with closed-response questionnaires: the narrow range of possible answers. The researcher may overlook potentially meaningful responses and not include
them among the range of possibilities. Closed-response questionnaires tend to be less exploratory in nature as they generally result in findings the researcher is expecting. The ability to encounter unexpected responses is almost completely eliminated. Finally, clear, concise closed-response items are difficult to write.

Although Dorynei (2003) disagrees with those who claim that all questionnaire data are invalid and unreliable, he does outline several potential problems that could influence the results from questionnaire data. He lists the following disadvantages: simplicity and superficiality of answers, unreliable and unmotivated respondents, respondent literacy problems, little or no opportunity to correct the respondents’ mistakes, social desirability (or prestige) bias, self-deception, acquiescence bias, halo effect, and fatigue effects.

Given the scope and objectives of this study a closed-response format was chosen in spite of the potential pitfalls inherent in this format. The closed-response format was chosen as the preferred method of data collection primarily due to the ability of quantitative data to facilitate direct teacher-student comparisons on a large scale. Although the inclusion of qualitative data might have complemented the quantitative data the large scale of the study made the gathering of meaningful qualitative data prohibitive. The main goal of this research was to make direct comparisons between the responses of 49 language teachers and more than 1400 students in an attempt to identify meaningful trends on the 20 + target items included. As such, it appeared that to be able to make any reasonably insightful and generalizable conclusions based on qualitative data, a separate study with different goals would need to be undertaken. That is to say that to directly
represent qualitative teacher-student differences with 1400 students and their 49 teachers across two questionnaires with over 20 items on each seemed unfeasible, or at least, not within the scope of this study. In depth, small-scale, case studies of 6-10 classes might lend themselves more to open-ended responses and even teacher and student interviews that would shed light on teacher-student differences within the sample population. In this scenario, teachers and students could be interviewed and their responses compared to provide a detailed qualitative comparison between each group’s perspective on the target issues in FL teaching. This researcher felt that given the scope and objectives of this study the inclusion of a qualitative component would have appeared somewhat forced and introduced to the research design without contributing meaningfully to the quantitative results. The inevitably large numerical discrepancy between the quantitative data and the qualitative data would make it difficult to relate one to the other.

PILOT TESTING

The original version of the first questionnaire, Effective Foreign Language Teacher (see Appendix A) was piloted on two separate occasions with two distinct sets of classes that were not part of the primary study. Teachers and students from five different classes were asked to carefully respond to the first version of the questionnaire and to offer any comments or feedback regarding the formatting of the questionnaire, the wording of the instructions and the individual items, and any other changes that they felt would be beneficial. The first version of this questionnaire contained 30 items and included the stem “Effective foreign language teachers should . . .” After receiving feedback from participating teachers and students and analyzing the response patterns,
revisions were made and the 30 items were pared down to 24. For example, those items that resulted in 100% agreement or disagreement among participants on the pilot instruments were altered or deleted so as to provide more valuable, interpretable responses. The same format was employed with a 4-point Likert scale and numbered statements following the same stem.

The second pilot administration of the questionnaire was performed in two separate beginning-level foreign language courses. A factor analysis of the initial 30-item questionnaire revealed 10 factors emerging from a 30-item questionnaire with some of the latent variables demonstrating only marginal loadings on one particular factor, i.e., .35 out of a total 1. It was clear that, indeed, the questionnaires for this study, although relatively short, included many different issues relative to FL teaching, a topic discussed further below in the section on instrumentation development. This determination was made due to the diverse nature of items from seemingly similar categories. That is to say that several items on grammar teaching may not demonstrate clear loadings on the same ‘grammar teaching’ factor with high correlations to other such items since they focus on particular aspects of grammar teaching, i.e., inductive vs. deductive, meaningful vs. mechanical, rather than grammar teaching as one construct. If the questionnaire were to contain factors based on a multi-item scale wherein several items regarding one particular issue of grammar teaching, e.g., inductive grammar, could be included, the factor analysis would have been more informative in shaping the current questionnaire. Hence, the results from the factor analysis did not serve to alter the questions on the questionnaire as much as a simple analysis of patterns in students’ responses. For examples, items with
categorical agreement or disagreement from all participants were either deleted or altered to provide more meaningful information. Likewise, the researcher analyzed students written comments on the questionnaire and revisions were made again to the questionnaire.

The idea for the second questionnaire (Appendix F and Appendix G) was to allow students to assess their teachers’ practices and for teachers to self-evaluate on those same practices included on the first questionnaire. Early drafts of the second questionnaire asked students and teachers to indicate the frequency of certain behaviors with the assumption being that the frequency of certain behaviors performed by a teacher could be interpreted as an evaluation of the teacher’s performance. This logic was subsequently judged faulty given that even though students may observe something frequently that the field at large agrees on as effective, they may feel that it is performed ineffectively. In essence, this approach would not allow students to subjectively evaluate whether they felt the teacher’s performance was effective, regardless of frequency. For this reason the “How effective?” column was added to allow for a true evaluation of teachers’ performance by students. The second questionnaire with both the “How often?” and “How effective?” columns contained 20 items and was piloted with two classes taught by the same teacher.

INSTRUMENTATION DEVELOPMENT AND CONTENT

Early drafts of the instrumentation paralleled Reber’s (2001) 80-item instrument used in her dissertation research. Reber’s instrument resulted from thorough and extensive research into current considerations regarding effective L2 teaching and
provided a point of departure for the creation of the Evaluation instrument pertinent to the present study. She includes nine overarching categories in her 80-item instrument: 1) Standards for Foreign Language Learning, 2) corrective feedback, 3) teachers and teacher behaviors related to communicative approaches, 4) focus on form in classroom SLA, 5) individual learner differences in FL learning, 6) strategies for foreign language learning, 7) theories about SLA, 8) teacher qualifications, and 9) assessment in foreign language teaching. As mentioned, the current instruments evolved after several revisions and after being piloted three times with different beginning level foreign language students. The breadth and depth of Reber’s instrument reflect her target population—experienced ACTFL-affiliated foreign language teachers. Presumably, most of these teachers would be familiar with technical jargon in the field of foreign and second language pedagogy. Given the context of the current study, Reber’s instrument proved much too lengthy, technical, and theoretical to be used for the purposes of this research which included students’ perspectives.

Unlike Reber’s study, the driving force behind the present research is to directly compare both teachers’ and students’ perspectives on foreign language teaching practices. Key to this comparison is the ability to present both groups with similarly worded instruments. Therefore, the following constraints on the instrumentation were imposed by this study’s research design: 1) in order to create an instrument whose items would be intelligible to all participants, i.e., beginning level students and experienced teachers, extra effort was made to reduce the amount of discipline-specific jargon that would not be accessible to those with little to no exposure to the field of second language
pedagogy; furthermore, some items included a short example to help disambiguate the target method or technique; 2) since it had been decided to administer the questionnaires during class time to increase response rate, it was necessary to reduce the number of items to 24 on the effective teacher questionnaire and 21 items on the frequency and evaluation instruments; 3) as this research attempts to discern teachers’ and students’ perspectives on concrete teaching practices within the teacher’s control, items pertaining to theoretical issues in SLA and second language pedagogy were omitted; 4) many participating departments create assessment instruments for multi-section classes, especially exams, and for this reason testing and assessment items were limited on the Evaluation Questionnaire to those that would reside within each teacher’s jurisdiction.

Several other considerations that influenced the instruments’ construction relate to negatively worded items, category titles and organization, and range of the Likert-scale. It was determined that including several negatively worded items would help reduce the ‘halo’ effect whereby participants tend to mark positive responses on the same side of the scale with little to no reflection. Similarly, given the abbreviated nature of the instrument and to avoid influencing participants’ responses, items were not divided into categories with titles or grouped together as some categories had fewer items than others. This was done to keep students from focusing on exactly how many items were devoted to each area, e.g., grammar or culture, and attempting to infer the importance of the category from that number. Moreover, this strategy attempted to keep the use of explicit, isolated category titles such as ‘grammar’ or ‘technology’ from evoking certain biases or prejudices. Finally, the decision was made to use a 4-point, or ‘forced choice’, Likert
scale with no middle or neutral option. The purpose of the 4-point scale versus an odd-numbered scale is to encourage students to introspect and consider more thoroughly each item and whether they agree or disagree with it. This approach still allows for the generation of class means on each item that may lie between two whole numbers so that a score of 2.5 on a 4-point scale may indicate, mathematically, a neutral or undecided reaction to a particular item from a group perspective. One drawback of the 4-point forced choice option is that students who truly did reflect on an item and remained undecided were not able to express their neutrality. In theory, some of those students would mark agree while others would mark disagree.

In order to provide the reader with a detailed explanation of the instrumentation, each questionnaire and its items will be outlined in this section. The first questionnaire for both teachers and students entitled “The Effective Foreign Language Teacher,”—hereafter denominated “Effective Teacher”—provided a common stem for all 24 Likert-scale statements with a 4-point scale ranging from “Strongly Agree” to “Strongly Disagree.” The stem, “Effective foreign language teachers should . . .”, was followed by 24 numbered statements that completed the sentence (see Appendix A). The instructions at the top of the page encouraged students to “reflect on your personal beliefs regarding what characterizes effective foreign language teaching.” Furthermore, within the instruction paragraph students were assured that “There are no right or wrong answers, just those that are right for you.”

In addition to the original factor analysis run on the 30-item version of the Effective Teacher questionnaire, an additional factor analysis was run on the final 24-
item version using all participating students’ and teachers’ responses. From a logical, content-based standpoint, seven general category groups appear to describe the nature of the items: Grammar Teaching, Error Correction, Target Language Use, Culture, Computer-Based Technology, Communicative Language Teaching Strategies, and Assessment. A factor analysis with a Varimax rotation was conducted generating seven factors, or latent variables, with a minimum Eigen value of 1.0 and it, too, generated seven factors or latent variables. Due to the myriad issues relative to FL teaching included on the Effective Teacher questionnaire and the complex nature of certain items, it was not surprising that the seven latent variables with an Eigen value over 1.0 explained just 50% of the variance and, in some cases, presented an unclear mathematical picture of what each factor might represent.

The first factor contained items relative to the use of both the target language and English by both students and teachers. Factor 2 appeared to represent student-to-student interaction as it included items regarding group and pair work and information gap activities. Factor 3 did not result in a clear, common variable as it included items from a wide range of topics such as task-based teaching (11), TPR (12), inductive grammar (20), and use of real-life materials instead of textbooks (21). Factor 4 presented a much clearer picture of the underlying latent variable, culture, after loading unequivocally on Items 3 and 9, both of which relate directly to culture. Factor 5 represents the tightest fit with the logical categories assigned by the researcher in that all three items identified by the factor analysis were the same three assigned to the category “Error Correction.” Factors 6 and 7 did not result in easily identifiable latent variables. Factor 6 included Items 18 and 19; the
former relating to the use of grammar in real-world contexts, the latter having to do with
native-like control of the FL by the teacher. Likewise, Factor 7 included seemingly
unrelated items, one relative to computer-based technology (1), and the other having to
do with focusing on grammar instead of information exchange.

From a psychometric standpoint, the difference in the mathematically generated
factors and the logical, topic-based categories can be explained in various ways. First, as
explained previously, the many different facets of a large, overarching category such as
Grammar Teaching might not lead to predictable, mathematical relationships in response
patterns. Participants’ responses to items concerning inductive versus deductive grammar
teaching might not load on the same factor as they are two separate, very distinct
approaches to teaching grammar, although both are clearly grammar teaching techniques.
Second, some items were multidimensional, intentionally pitting one approach to
language teaching against another within the same item. For example, Item 16 states
“Effective foreign language teachers should mostly use activities that practice specific
grammar points rather than activities whose goal is to merely exchange information.”
This item contrasts practicing grammar with more communicative approaches that value
information exchange and was worded in this manner to purposely spur reactions from
students and teachers on two salient issues in FL teaching: grammar practice versus
communication practice. The third possibility as to the unclear factor loadings might be
the use of double-barreled items where clauses or phrases are conjoined by “and” or “or”
and, thus, include more than one variable per item. As explained elsewhere, this
instrument attempts to cover a wide range of FL issues with only 24 items due to the time
restrictions imposed by the research design. Notwithstanding the unclear statistical picture painted by the factor analysis, the primary concern of the current research was to compare teachers’ and students’ reactions to the same items so that teachers and students were exposed to the exact same wording. As a result, the logical, topic-centered categories assigned to each item by the researcher will be used hereafter in discussing common, underlying themes between items. Table 4 presents a summary of the factor loadings from a Varimax rotation on all 24 items on the Effective Teaching Questionnaire. Table 5 contains a summary of the questionnaire’s items and how they fit into seven different logical, content-based pedagogical categories.

Table 2. Summary of Factor Loadings on Items on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Item Number (Factor Loading)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1—Language Use</td>
<td>4 (.46), 6 (-.60), 7 (.72), 14 (.64), 17(-.62), 22 (.4)</td>
</tr>
<tr>
<td>Factor 2—Group Work/Interaction w/ Classmates</td>
<td>2 (.66), 15 (-.59), 23 (.63), 24 (.63)</td>
</tr>
<tr>
<td>Factor 3—?</td>
<td>11 (.63), 12 (.31), 20 (.60), 21 (.66)</td>
</tr>
<tr>
<td>Factor 4—Culture</td>
<td>3 (.81), 9 (.83)</td>
</tr>
<tr>
<td>Factor 5—Error Correction</td>
<td>5 (.77), 8 (.42), 13 (-.66)</td>
</tr>
<tr>
<td>Factor 6—?</td>
<td>18 (.77), 19 (.47)</td>
</tr>
<tr>
<td>Factor 7—?</td>
<td>1 (.44), 16 (.68)</td>
</tr>
</tbody>
</table>
Table 3. Summary of Item Distribution across Logical Categories on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Category</th>
<th>Item Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Teaching</td>
<td>10, 16, 18, 20</td>
</tr>
<tr>
<td>Error Correction</td>
<td>5, 8, 13</td>
</tr>
<tr>
<td>Target Language Use</td>
<td>7, 14, 17, 19, 22</td>
</tr>
<tr>
<td>Culture</td>
<td>3, 9, 21</td>
</tr>
<tr>
<td>Computer-based Technology</td>
<td>1</td>
</tr>
<tr>
<td>Communicative Language Teaching Strategies</td>
<td>2, 4, 11, 12, 15, 21, 23, 24</td>
</tr>
<tr>
<td>Assessment</td>
<td>2, 6, 10, 23</td>
</tr>
</tbody>
</table>

In order to illustrate how each logical category was manifested in the questionnaire’s items, each individual item will be reproduced organized by section with the corresponding number of the item retained. Each item was preceded by the stem “Effective foreign language teachers should . . .”

**Category: Grammar Teaching**

10-**not** grade language production (i.e. speaking and writing) primarily for grammatical accuracy.

16-mostly use activities that practice specific grammar points rather than activities whose goal is to merely exchange information.

18-**not** present a particular grammar point without illustrating how the structure is used in a specific, real-world context.
20-teach grammar by giving examples of grammatical structures **before** explaining the grammar rules.

**Category:** Error Correction

5- **not** correct students immediately after they make a mistake in speaking.

8-only correct students indirectly when they produce oral errors instead of directly, e.g. correctly repeating back to them rather than directly stating that they are incorrect.

13-address errors by immediately providing explanations as to why students’ responses are incorrect.

**Category:** Target Language Use

7- **not** use English in the foreign language classroom.

14-require students to speak in the foreign language beginning the first day of class.

17-ask students to begin speaking the foreign language only when they feel they are ready to.

19-speak the foreign language with native-like control of both grammar and accent.

22- **not** simplify or alter how they speak so that students can understand every word begin said.

**Category:** Culture

3-devote as much time to the teaching of culture as to the teaching of language.
9-be as knowledgeable about the culture(s) of those who speak the language as the language itself.

21-use predominantly real-life materials (e.g. music, pictures, foods, clothing) in teaching both the language and the culture rather than the textbook.

**Category:** Computer-based Technology

1-frequently use computer-based technologies (internet, CD-ROM, email) in teaching the language.

**Category:** Communicative Language Teaching Strategies

2-base at least some part of students’ grades on completion of assigned group tasks.

4-require students to use the language outside of class with other speakers of the language (e.g. internet, email, clubs, community events, etc.)

11-teach the language primarily by having students complete specific tasks (e.g. finding out prices of rooms and rates at a hotel) rather than grammar-focused exercises.

12-have students respond to commands physically in the foreign language (e.g. “stand up,” “pick up your book,” etc.)

15-not use predominantly small groups or pair work to complete activities in class.

21-use predominantly real-life materials (e.g., music, pictures, foods, clothing) in teaching both the language and the culture rather than the textbook.
23-base at least some part of students’ grades on their ability to interact with classmates successfully in the foreign language.

24-use activities where students have to find out unknown information from classmates using the foreign language.

**Category:** Assessment

2-base at least some part of students’ grades on completion of assigned group tasks.

6-allow students to respond to test questions in listening and reading via English rather than the foreign language.

10-*not* grade language production (i.e., speaking and writing) primarily for grammatical accuracy.

23-base at least some part of students’ grades on their ability to interact with classmates successfully in the foreign language.

Attached to the Effective Teacher questionnaire was a single-sided demographic survey entitled *Student Information* (see Appendix B). The student information sheet included detailed demographic questions and was intended to gather student data that could be revealing when analyzed as independent variables of each questionnaire and each item. It was divided into three sections: *Personal Data, Current Academic Data, Foreign/Second Language Experience*. The *Personal Data* section included two questions, one regarding gender and the other asking age. As all questionnaires were administered on scantron sheets it was necessary to provide different age groups, i.e., 18-20, 21-23, 24-27, 28-35, 36-50, 50+. The *Current Academic Data* portion included five
questions including questions regarding current status as a student, overall GPA, reason for taking class, level of perceived difficulty of the class, and expected grade from class. The section on students’ foreign and second language learning experience included four questions relative to whether students might be classified as heritage language learners. The remaining five questions queried students’ self-assessed proficiency, exposure to formal education in the language of study or another language, stays abroad in locations where the language of study was spoken, and quality of their overall experience with foreign language learning. (See Appendix C for results from Student Information Questionnaire.) In order to provide the reader with a concrete idea of the categories and questions included in this questionnaire, they are reproduced here:

**Personal Data:**

Q1: What is your gender?

Q2: What is your age?

**Current Academic Data:**

Q3: Which of the following best describes your status as a student at the university?

Q4: What is your approximate, overall GPA at the university?

Q5: Why are you taking this class?

Q6: How difficult do you find this specific foreign language class to be?

Q7: Truthfully, what grade do you expect from this particular class?

**Foreign/Second Language experience:**

Q8: Did you hear the language of this class frequently as a child?
Q9: Did you speak the language of this class frequently as a child?

Q10: As children, did your parents or grandparents speak and hear the language of this class?

Q11: How old were you when you started learning English?

Q12: Which best describes your self-assessed language proficiency in the language of this class?

Q13: How much formal education have you had in the language of this class?

Q14: How long have you lived abroad consecutively in a country where the language of this class is spoken?

Q15: Have you studied another language other than this language and your native language?

Q16: How would you describe your overall experience with classroom foreign language learning?

Similar to the students, teachers were asked to fill out a demographic survey (see Appendix D for Teacher Information Questionnaire and Appendix E for summary of teachers’ responses) attached to the Effective Teacher questionnaire to allow for analyses of demographic variables. Their questionnaire included the following sections: Personal Data, Current Academic Status, Training, Previous Language Teaching Experience, and Current Teaching Responsibilities. In contrast with the student information questionnaire, the teacher information sheet grouped gender and age with questions regarding exposure to the language and culture, proficiency in the language, and years of domestic and/or foreign residence. The following section, Current Academic Status, requested that
instructors identify the category that best described their standing at the university, e.g., graduate student in same department. Instructors’ pedagogical training was determined with one question asking how many teacher training courses instructors had taken. The section on previous language teaching experience was comprised of five questions concerning the amount and type of teachers’ experience with language teaching. The last section, Current Teaching Responsibilities, simply asked whether teachers had taught the class they were currently teaching before. The five categories with accompanying questions are included verbatim to concretize the previous discussion.

**Personal Data:**

Q1: What is your gender

Q2: What is your age

Q3: As a child, did you frequently hear the language you teach?

Q4: As a child, did you frequently speak the language you teach?

Q5: As children, did your parents or grandparents hear and speak the language you teach?

Q6: Which term best defines your language proficiency in the language you teach?

Q7: Which terms best defines your familiarity with the culture(s) affiliated with the languages you teach?

Q8: How long have you lived abroad consecutively in a country where the language of this class is spoken?

Q9: If you are not from the U.S., how long have you resided in the U.S.?
Current Academic Status:

Q10: Which of the following best describes your current status at the university?

Training:

Q11: How many teacher training courses have you formally taken (e.g. teaching methods, testing)?

Previous Language Teaching Experience:

Q12: Overall, how many years have you been teaching language-related courses?
Q13: How many years have you been teaching language-related courses in the U.S.?

Which of the following levels and course types have you taught?
Q14: General Level
Q15: Course Level
Q16: Course Type

Current Teaching Responsibilities:

Q17: Have you taught the class you are currently teaching before?

The aforementioned Effective Teacher, Student Information, and Teacher Information questionnaires, were administered during the first classroom visit (weeks 4-6 of semester) while the second classroom visit (weeks 10-15 of semester) focused on perceptions of teaching behavior frequency and effectiveness. This questionnaire, hereafter called the Evaluation questionnaire (see Appendix F and Appendix G) was designed to allow students and teachers, first, to reflect on how often they perceived certain teaching practices to be taking place in their classrooms, and, second, to evaluate
how effective the teacher was in performing those practices. The design of this questionnaire differed from the first in that students were presented with 21 statements that completed the stem, “In this class, my foreign language teacher . . .” At the end of each statement were three columns: the first column offered a 4-point Likert-scale that included, from left to right, F (frequently), S (sometimes), R (rarely), N (never); the second column to the right of the first was another 4-point Likert scale made up of VE (very effective), E (effective), LE (limited effectiveness), IE (ineffective); the final column had only one option which was NA (not applicable). The Evaluation Questionnaire requested that students make a judgement of how often their teacher performed certain skills and how effective the teachers were in performing them. Hence, above the first 4-point Likert-scale column was “How often?” and above the second was “How effective?” The instructions asked that students consider “1) how often your teacher performs each behavior in your class and 2) when the behavior does take place, how effectively your teacher performs it . . .” If students indicated that a behavior “Never” occurred they were asked to “please mark ‘NA.’” Apparently, some students from the total number participating misunderstood this last procedure and mistakenly marked a degree of effectiveness after marking “Never.” In cases where this took place the researcher adjusted the data so that when “Never” was bubbled, the effectiveness column was left blank and the “NA” column was marked.

The teachers’ version of the Evaluation Questionnaire included similar directions and the same items with the stem being changed to state the following: “As a foreign language teacher in my current class(es) of the same level, I . . .” Accordingly, the verbs
of each of the 21 items were inflected correctly for first person singular. The Evaluation Questionnaire paralleled closely the Effective Teacher questionnaire with several changes. As mentioned previously, many of the larger foreign language departments do not allow individual instructors to create their own tests or graded assignments. For this reason some questions had to be omitted from the Evaluation Questionnaire so as to reflect behaviors and decisions within the teachers’ scope of influence. Additionally, the Evaluation Questionnaire did not include negatively-worded items as it was illogical to ask participants how often a teacher performed something and how effective they were in performing it if it the item was stated using “not”. An example from the Effective Teacher questionnaire might help elucidate this point. Item 15 states that “Effective foreign language teachers should not use predominantly small groups of pair work to complete activities in class.” On the Evaluation Questionnaire it would be illogical and incoherent to ask students and teachers to reflect on how often and how effective teachers “do not use predominantly small groups or pair work to complete activities in class”. On the Evaluation Questionnaire this item was worded in the positive in the following manner: “In this class, my foreign language teacher uses small-group work to complete in-class activities in the foreign language.” Therefore, the threat of a ‘halo’ effect was regarded as a non-issue on the Evaluation Questionnaire. Finally, certain items from the Effective Teacher questionnaire needed to be re-worded to better allow for a graded observation and evaluation rather than a simple agree or disagree. For example, one of the target language use items on the Effective Teacher instrument states that “Effective Foreign Language teachers should … ask students to begin speaking the foreign language
only when they feel they are ready to.” This item was reworded for the Evaluation Questionnaire to read “In this class my foreign language teacher/As a foreign language teacher in my current class(es) of the same level, I . . . encourages/encourage students to speak the foreign language within the classroom with others.” Items such as these demonstrate the awkwardness of asking students to evaluate the statement “In this class my foreign language teacher asks students to begin speaking the foreign language only when they are ready to.” Table 6 summarizes the distribution of items across categories on the Evaluation Questionnaire.

Table 4. Summary of Item Distribution across Categories on Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>Category</th>
<th>Item Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Teaching</td>
<td>7, 14, 16</td>
</tr>
<tr>
<td>Error Correction</td>
<td>4, 8, 10</td>
</tr>
<tr>
<td>Target Language Use</td>
<td>5, 11, 15, 18, 21</td>
</tr>
<tr>
<td>Culture</td>
<td>2, 17</td>
</tr>
<tr>
<td>Computer-based Technology</td>
<td>1</td>
</tr>
<tr>
<td>Communicative Language Teaching Strategies (i.e., Task-based Language Teaching, TPR, Interaction/Negotiation of Meaning)</td>
<td>6, 9, 12, 13, 17, 19, 20</td>
</tr>
<tr>
<td>Assessment</td>
<td>7, 20</td>
</tr>
</tbody>
</table>

The items as they appeared on the questionnaire are presented below by category with the stem removed. The stem for the students’ version stated “In this class, my foreign language teacher . . .” while the teachers’ version stated “As a foreign language teacher in my current class(es) of the same level, I . . .” For convenience, the third-person
singular ‘s’ is added in parentheses next to the main verb of each item below to indicate that the students’ version matched the stem mentioned above.

**Category**: Grammar Teaching

7-grade(s) exercises in the foreign language (i.e. speaking, writing) for grammatical accuracy.

14-present(s) lessons that emphasize particular grammar aspects by illustrating how the structure is used in a real-world context (e.g. the future tense in setting goals or New Year’s resolutions).

16-teach(es) grammar by explaining the grammatical rules **before** giving concrete examples for students to consider.

**Category**: Error Correction

4-correct(s) students **directly** when they make oral mistakes (e.g. by specifically stating that a certain structure they have used is incorrect).

8-correct(s) students indirectly when they make oral errors (e.g. by correctly repeating back to them, pausing after the error, asking them to repeat what they said, etc.)

10-address(es) errors by providing explanations as to why students’ language is incorrect.

**Category**: Target Language Use

5-make(s) use of English in the foreign language classroom.

11-encourage(s) students to speak the foreign language within the classroom with others.
15-demonstrate(s) native-like control in expressing myself (her/himself) in the foreign language (i.e. use accurate grammar and native-like accent).

18-alter(s) how I (she/he) speak(s) so that students can understand what is being said.

21-use(s) the foreign language in the classroom.

**Category**: Culture

2-devote(s) time to the teaching of culture(s) of those who speak the language.

17-use(s) real-life materials (e.g. music, pictures, food, clothing) in the foreign language classroom.

**Category**: Computer-based Technology

1-use(s) computer-based technologies other than word processing (e.g. internet, CD-ROM, email) in teaching the language.

**Category**: Communicative Language Teaching Strategies

6-have (s) students complete tasks with a real-life purpose in the foreign language (e.g. planning a vacation using the internet or travel brochures.).

9-have (s) students respond to commands physically in the foreign language (e.g. “stand up,” “pick up your book,” etc.).

12-use(s) small-group work to complete in-class activities in the foreign language classroom.
13-use(s) activities whose goal is to exchange information rather than practicing specific grammar points (e.g. talking about an individual’s daily activities vs. conjugating verbs in table form).

17-use(s) real-life materials (e.g. music, pictures, food, clothing) in the foreign language classroom.

19-use(s) activities where students have to find out information from classmates using the foreign language.

20-grade(s) activities where students must interact with classmates in the foreign language (e.g. in pairs or small groups to complete role plays, dialogues, presentations, etc.).

Category: Assessment

7-grade(s) exercises in the foreign language (i.e. speaking, writing) for grammatical accuracy.

20-grade(s) activities where students must interact with classmates in the foreign language (e.g. in pairs or small groups to complete role plays, dialogues, presentations, etc.)

Clearly, the categories and items from each questionnaire do not provide a broad or in-depth coverage of issues currently under scrutiny in the field of foreign language pedagogy. However, the items do address many issues that have relevance in today’s foreign language classrooms. Moreover, the large-scale on which the study was conducted and the detailed nature of the teacher-student comparisons make the data
powerful in assessing the match between teachers’ and students’ perceptions at the target institution—the University of Arizona.

The following table provides a concordance between each item from the Effective Teacher and Evaluation Questionnaires and relevant references that demonstrate the salience of each issue in foreign language pedagogy. These principles and constructs were reviewed in Chapter 2.

Table 5. Item and Abbreviated Reference Concordance.

<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Category</th>
<th>Effective Teacher Questionnaire</th>
<th>Evaluation Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Topic</td>
<td>Pages</td>
<td>Theme</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Anton &amp; DiCamilla, 1999; Cook, 2001; Duff &amp; Polio, 1990; Gass, 1997;</td>
<td>Target Language Use</td>
<td>7, 14, 17, 19, 22, 23</td>
<td>5, 11, 15, 18, 20, 21</td>
</tr>
<tr>
<td>Gass &amp; Varonis, 1994; Krashen, 1982; Krashen &amp; Terrell, 1983; Levine,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003; Long, 1996; Polio &amp; Duff, 1994; VanPatten, 1996</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTFL Standards for Foreign Language Learning, 1999; Brooks, 1968;</td>
<td>Culture</td>
<td>3, 9, 21</td>
<td>2, 17</td>
</tr>
<tr>
<td>Lange, 1999; Sapir, 1949; Seelye, 1997; Valdes (Ed.), 1986; Whorf, 1956</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beauvois, 1998; Blake, 1998, 2001; Bush, &amp; Terry (Eds.), 1997; Kern,</td>
<td>Computer-based Technology</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1995b; Kern &amp; Warschauer, 2000; Muyskens, J. (Ed.), 1997; Payne &amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whitney, 2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asher, 1972, 2003; Asher, Kusudo, de la Torre, 1974; Ellis, 1999, 2003;</td>
<td>Communicative Language</td>
<td>11; 12; 2, 4, 15, 21, 23, 24</td>
<td>6; 9; 12, 13, 17, 19, 20</td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ballman, 1998; Bachman, 1990; Bachman &amp; Palmer, 1996; Bachman &amp; Cohen,</td>
<td>Assessment</td>
<td>2, 6, 10, 23</td>
<td>7, 20</td>
</tr>
<tr>
<td>1998; Bernhardt, 1991; Ekbatani &amp; Pierson, 2000; Hammerly, 1991;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PROCEDURES

Within the first four to six weeks of Spring semester 2005, the researcher or a research assistant entered the 83 classes and participating students and teachers filled out two questionnaires in the following order: 1) a 24-item, Likert-scale questionnaire regarding perceptions of effective foreign language teachers, i.e., Effective Teacher Questionnaire (see Appendix A), and 2) a demographic questionnaire (see Appendix B for Student Information Questionnaire and Appendix D for Teacher Information Questionnaire). The Effective Teacher Questionnaire was used to answer Research Questions 1-3, which assessed students’ and teachers’ perceptions overall with Research Question 3 directly comparing teachers’ and students’ perceptions, and Research Question 4 which made individual comparisons between teachers and their students. The rationale behind the administration of the questionnaire within weeks 4-6 was to minimize the tendency of students to reflect on their current teachers’ practices when answering the effectiveness questionnaire as well as to reduce the impact of students who were still adding, dropping, and changing classes before the deadline for authorized drops. Teachers who taught more than one section or level of a language were only required to fill out one Effective Teacher questionnaire as the questions were global and not applicable to one individual class per se. Only four of the 49 instructors taught more than one level.
Between weeks 10 and 15 in the semester, students filled out the questionnaire asking for perceptions of the frequency of target teaching behaviors and their evaluation of their teachers’ performance of those behaviors. This questionnaire addressed Research Questions 5, and 6, which compare teachers’ and students’ perceptions of the frequency of occurrence of teaching behaviors and teachers effectiveness in implementing them, and used in conjunction with class level and language type will allow Research Question 7 to be answered—the effect of language type and level on students’ responses on all questionnaires. Students also responded to six questions taken from the university’s Teacher-Course Evaluation (TCE) whose responses will answer Research Question 8, comparison of students’ responses on selected TCE questions and the evaluation column of the Evaluation Questionnaire, and Research Question 9, an analysis of the relationship between students’ responses on all questionnaires and students’ TCE responses. During that same visit teachers completed the same questionnaire worded in the first person asking for their perceptions of the frequency of certain teaching behaviors and their self-evaluations of those practices. The timing for this second visit reflected the need for both groups to have experienced a sufficient amount of instruction to make a valid observation of the frequency of teaching behaviors and a reliable assessment of teaching throughout the semester.

Students who were absent during the administration of the two questionnaires during the first classroom visit were given the opportunity to fill out questionnaires during the second visit. Students who missed the first class visit were asked to also fill out a demographic questionnaire during the second visit so their unique demographics
would be known. Each teacher and student was asked to produce a unique alpha-numeric identifier that was comprised of their 4-digit birthday without the year and the first three letters of their mother’s first name, e.g., 0123ABC. The strategy behind this identifier was to avoid giving the students a random number that they might not remember or misplace from the first to the second visit. By using the alphanumeric code described above, students would not need to memorize or keep track of a meaningless code. As the comparisons were between teachers and whole classes, or all teachers and all classes, data from students who attended the first class and were absent for the second visit, or vice versa were included.

DATA ANALYSIS

Students’ and teachers’ responses on the Effective Teacher questionnaire and the Evaluation Questionnaires were subjected to statistical analyses that generated descriptive statistics, Pearson product-moment correlation coefficients, and inferential statistics, i.e., t-tests for means comparisons, and Analysis of Variance (ANOVA) for effect of language type and class level on students’ responses to each questionnaire. A multiple regression analysis was used to assess the relationship between students’ responses to each questionnaire and their ratings of their teachers on the TCE questions.

All teachers’ and all students’ responses on the Effective Teacher instrument were analyzed separately per Research Question 1 and Research Question 2 using descriptive statistics. Then, all teachers’ responses were compared to all students’ per item as dictated by Research Question 3 which required inferential statistics (t-tests). Whenever a large number of t-tests were run on the same questionnaire, a Bonferroni
adjustment procedure was performed in interpreting all results from inferential statistics.
The traditional alpha level of .05 used in the social sciences was divided by the number
of t-tests, or ANOVA, run on respective questionnaires—24 for the Effective Teaching
questionnaire and 21 for the Evaluation Questionnaire—to produce adjusted values of
.0021 and .0024 respectively as new alpha levels. When all students are compared to all
teachers, the Bonferroni adjustment helps avoid Type I errors where significant \( p \) values
result by chance due to the large number of analyses that are run, 24 in the case of the t-
tests run on responses from the Effective Teacher questionnaire and 21 in the case of the
ANOVA analyses run on responses from the Evaluation Questionnaire.

Research Questions 4, 5, and 6 compared each teacher’s responses to those of
his/her classes on all three facets of both questionnaires, i.e., ideals of effective foreign
language teachers, frequency of occurrence of their teachers’ behaviors, and the
effectiveness of their teachers’ performance of those practices—all of which required
Pearson product-moment correlations as well as descriptive statistics. Comparisons
between individual teacher’s raw scores and their students in the form of t-tests were not
a possibility on the Effective Teacher questionnaire as there would be ‘0’ degrees of
freedom and hence no variance. An additional reason for the impossibility of directly
comparing individual teacher’s responses to students’ on Research Questions 5 and 6
(Evaluation Questionnaire) was the fact that the recasting of the stem to match teachers’
and students’ perspectives creates two dependent variables rather than one. Although
students and teachers were asked to reflect on the same individual’s teaching behaviors,
the difference in perspective represented by different prompts, i.e., “My teacher . . .” v. “I
generates, two distinct, dependent measures. Therefore, in comparing all students’
and all teachers’ perceptions overall of frequency of occurrence of certain behaviors and
perceived effectiveness of teachers’ performance, a Pearson product-moment correlation
was run. For individual teacher comparisons, descriptive statistics were used to identify
common trends as well as noteworthy outliers. These descriptive statistics were
calculated by taking the difference between each individual teacher’s response value on
each item and subtracting it directly from his/her students’ mean response on the same
item. Then, an average was taken across all responding teachers and their individual
teacher-student difference per item to give an overall averaged teacher-student difference
for each item.

Research Question 7 involves the examining of the relationship between
independent variables, namely class language type (commonly taught languages versus
less commonly taught languages) and level (first and second year), and their relationship
to the dependent variable—students’ responses on each questionnaire. This analysis
required an inferential statistic, ANOVA, to surmise the effect of each independent
variable, language type and level, on the dependent variable, students’ responses to the
Effective Teacher and Evaluation Questionnaires.

Research Question 8 compares students’ responses on selected TCE questions to
their responses on the Evaluation Questionnaire. In order to execute this comparison it
was necessary to use a Pearson product-moment correlation analysis which generates a
correlation coefficient (r) between two variables, namely, students’ responses on the
evaluative column of the Evaluation Questionnaire and their responses on the TCE questions.

The final research question, Research Question 9, pertains to the relationship between students’ responses on each questionnaire and the selected TCE questions. In order to answer this question a multiple regression analysis was conducted to determine the degree of correlation between students’ responses to the TCE questions and the Effective Teacher Questionnaire and the frequency and evaluation columns of the Evaluation Questionnaire. A correlation (r) between all variables included in the model was needed in order to perform the multiple regression analysis.

SUMMARY

An explanation of the methodology adopted in the current study was presented in Chapter 3 to facilitate the interpretation of the results and subsequent conclusions to be treated in Chapter 4. This chapter’s explanation detailed the nature of the study’s context and participants, the justification for the use of a quantitative questionnaire, the considerations that guided the design of the current instrumentation, and the ensuing statistical analyses to be performed. Chapter 4 will systematically answer each of the nine research questions through the statistical analyses performed on the data.
CHAPTER 4
DATA ANALYSIS AND RESULTS

INTRODUCTION

This chapter will describe the statistical analyses that were conducted on the data and the resulting findings. After a brief explanation and justification of the statistical tests that were run, each research question will be answered in turn. While Chapter 4 focuses more strictly on the nature of the statistical tests that were run and the corresponding results, some discussion will be included when appropriate. The research questions are reproduced here to facilitate the reading and comprehension of Chapter 4:

Research Question 1

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do students consider to be reflective of effective foreign language teachers?

Research Question 2

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do these students’ teachers consider to be reflective of effective foreign language teachers?

Research Question 3

Based on a similar 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), how do students’ and teachers’ perceptions of effective foreign language teachers coincide or differ?

Research Question 4
How do individual teachers’ perceptions of effective foreign language teachers coincide with or differ from those of their students’?

Research Question 5

How do students’ and teachers’ perceptions of the frequency of occurrence of certain teaching behaviors in their classrooms coincide or differ, both overall and in a teacher-by-teacher comparison on a 21-item, Likert-scale questionnaire (Evaluation Questionnaire)?

Research Question 6

How do teachers’ self-evaluations coincide with or differ from their students’ evaluations of them overall and individually?

Research Question 7

What impact do language type and class level exert on students’ responses to both the 24-item (Effective Teacher) and 21-item (Evaluation) questionnaires?

Research Question 8

How do students’ evaluations of their foreign language teachers on the 21-item language-specific instrument (Evaluation Questionnaire) coincide with or differ from their evaluation of their teachers on selected items from the general, university-wide evaluation form, i.e., the Teacher-Course Evaluation (TCE)?

Research Question 9

What relationship exists between students’ response patterns on both the Effective Teacher and Evaluation Questionnaires and students’ responses to selected items from the TCE instrument?
STATISTICAL PROCEDURES

The overriding purpose of this study is to achieve a detailed comparison between two distinct groups, namely, teachers and students, on the issues in foreign language teaching included on the study’s questionnaires. Therefore, statistical tests were performed to compare students’ and teachers’ responses on each questionnaire. In the case of the Effective Teacher questionnaire, two-tailed, independent group t-tests were used to compare teachers’ and students’ overall means since both teachers and students responded to the exact same wording on this questionnaire. As explained previously, on both the Effective Teacher and Evaluation Questionnaires individual teachers were compared to their students using descriptive and correlation analyses (Research Questions 4, 5, and 6). Direct means comparisons using inferential statistics, i.e., t-tests, were not considered to be an appropriate statistical analysis to use when comparing individual teachers and their students due to the lack of variance presented by ‘0’ degrees of freedom in the case of the teacher as well as the recasting of the stem on the Evaluation Questionnaire to match students’ and teachers’ perspectives.

Research Questions 1 and 2 provide an overview of teachers’ and students’ independent perceptions of the FL teaching issues examined and, thus, required only descriptive statistics. Research Question 3 compared the two groups overall, teachers and students, on the same questionnaire and allowed for a means comparison analysis to be conducted via a t-test for each item.

Research Questions 4, 5, and 6 focus on differences between individual teachers and their students. In order to provide a summary of these differences between individual
teachers and their students, an averaged teacher-student value across all teachers was calculated per item. Overall student means and teacher means are reported per item with accompanying Pearson product-moment correlations for each item. For each individual teacher a Pearson product-moment correlation was conducted between each individual teacher’s responses across all items of each questionnaire and column to produce a correlation coefficient for each teacher and his/her students on an entire questionnaire or column.

In the case of Research Question 7, it was necessary to calculate whether the impact of language type (less commonly taught v. commonly taught languages), level (first and second year), and the interaction between the two presented a statistically significant interaction effect with students’ responses on each questionnaire. For this reason, it was necessary to use an Analysis of Variance (ANOVA) procedure for each item on the questionnaires to answer Research Question 7.

Research Questions 8 and 9 include a comparison between students’ responses on the instrumentation used in the current study and students’ responses on questions taken from the official Teacher-Course Evaluation (TCE) used at the University of Arizona. Since various differences exist between the TCE and the current evaluation instrument, such as the use of a 5-point scale on the TCE instead of a 4-point scale used on the Evaluation Questionnaire and a more generalized set of constructs for effective teaching, correlation analyses were conducted. Of the six total questions taken from the TCE it was determined that only two of them related directly to the teacher and to teacher behaviors. Hence only these two were included in the statistical analyses conducted to answer
Research Questions 8 and 9. For Research Question 8 which compared two variables, namely, students’ responses to the evaluation column of the Evaluation Questionnaire and their responses to two specific TCE questions, a Pearson product-moment correlation was used. In the case of Research Question 9, students’ responses on each respective questionnaire—Effective Teacher, Evaluation, and TCE—were correlated using a Multiple Regression Analysis. Given the exploratory nature of these final two research questions all comparisons and correlations that resulted from the analyses must be regarded as tentative and preliminary. All of these statistical analyses were performed using the statistical computer software package SAS version 9.1.

RESEARCH QUESTION 1

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do students consider to be reflective of effective foreign language teachers?

Students most strongly agreed with the following items: 8 (only correct oral errors indirectly), 9 (be as knowledgeable about culture as language), 12 (have students respond to physical commands—TPR), 13 (address errors with immediate explanation), 18 (not present grammar without real-world context), 19 (speak with native-like control of language), 21 (use real-life materials in teaching lang. & cult.), and 24 (engage students in information gap activities). These were items that resulted in average responses between ‘Agree’ (3) and ‘Strongly Agree’ (4). The highest level of agreement among students was with Item 19, which states that effective foreign language teachers should have native-like control of grammar and accent. Those items that evoked more
moderate opinions from students, between a mathematically neutral 2.5 and a 3
(‘Agree’), included questions 1 (use computer-based technologies), 2 (assess group
tasks), 3 (devote time to culture), 4 (have students use language outside of class), 10 (not
grade production for accuracy), 14 (require students to speak FL first day of class), 16
(use activities to practice grammar points rather than information exchange), 20 (teach
grammar with examples before rules), 22 (not simplify speech), and 23 (assess successful
interaction with classmates). Students expressed disagreement with the following items
as they rated all of these questions less than 2.5: 5 (not correct immediately), 6 (assess
listening and reading via English), 7 (not use English), 11 (have students complete
specific tasks rather than grammar), 15 (not use small group or pair work), and 17 (ask
students to speak FL only when ready). The item that produced the lowest score (2.03)
was Item 17 which states “Effective foreign language teachers should ask students to
begin speaking the foreign language only when they feel they are ready to.” Apparently,
students expect teachers to ask them to begin speaking the language before the students
themselves feel they are ready or they desire to be pushed as learners. Table 6 presents
student means on each item of the Effective Teacher Questionnaire in rank order from
highest to lowest agreement with the question number in parentheses following the item.
In cases where the same mean resulted on separate items (e.g., Items 6 and 7) both items
were given the same rank resulting in less than 24 total rankings.
Table 6. Student Means on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Rank of Item-Item</th>
<th>“Effective foreign language teachers should . . .” Student Means (n = 1414) 1=SD; 4=SA</th>
<th>“Effective foreign language teachers should . . .” Student Means (n = 1414) 1=SD; 4=SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-related to native-like control of language (Q19)</td>
<td>3.25</td>
<td>13-use activities to practice grammar points rather than information exchange (Q16) 2.72</td>
</tr>
<tr>
<td>2-address errors with immediate explanation (Q13)</td>
<td>3.13</td>
<td>14-have students use language outside of class (Q4) 2.69</td>
</tr>
<tr>
<td>3-use real-life materials in teaching lang. &amp; cult. (Q21)</td>
<td>3.11</td>
<td>15-not grade production for accuracy (Q10) 2.66</td>
</tr>
<tr>
<td>4-be as knowledgeable about culture as language (Q9)</td>
<td>3.10</td>
<td>16-not simplify speech (Q22) 2.60</td>
</tr>
<tr>
<td>5-not present grammar without real-world context (Q18)</td>
<td>3.09</td>
<td>17-require students to speak FL first day of class (Q14) 2.55</td>
</tr>
<tr>
<td>6-have students respond to physical commands (TPR) (Q12)</td>
<td>3.06</td>
<td>18-use computer-based technologies (Q1) 2.53</td>
</tr>
<tr>
<td>7-engage students in information gap activities (Q24)</td>
<td>3.05</td>
<td>19-have students complete specific tasks rather than grammar (Q11) 2.46</td>
</tr>
<tr>
<td>8-only correct oral errors indirectly (Q8)</td>
<td>3.02</td>
<td>20-assess listening and reading via English (Q6) 2.36</td>
</tr>
<tr>
<td>9-assess group tasks (Q2)</td>
<td>2.97</td>
<td>20-not use English (Q7) 2.36</td>
</tr>
<tr>
<td>10-assess successful interaction with classmates (Q23)</td>
<td>2.95</td>
<td>21-not correct immediately (Q5) 2.12</td>
</tr>
<tr>
<td>11-teach grammar with examples before rules (Q20)</td>
<td>2.88</td>
<td>21-not use small group or pair work (Q15) 2.12</td>
</tr>
<tr>
<td>12-devote time to</td>
<td>2.74</td>
<td>22-ask students to speak 2.03</td>
</tr>
</tbody>
</table>
RESEARCH QUESTION 2

Based on a 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), what teaching behaviors do these students’ teachers consider to be reflective of effective foreign language teachers?

Teachers were most in agreement with Items 2 (assess group tasks), 9 (be as knowledgeable about culture as language), 12 (have students respond to physical commands—TPR), and 24 (engage students in information gap activities), all of which scored means of 3.3 or higher on a 4-point (1-4) scale. Item 24, relative to the use of information gap activities, was the most strongly agreed upon item by teachers. Other items that teachers agreed upon with scores greater than 3 were 3 (devote time to culture),
4 (have students use language outside of class), 5 (not correct immediately), 8 (only correct oral errors indirectly), 14 (require students to speak FL first day of class), 18 (not present grammar without real-world context), 19 (speak with native-like control of language), 21 (use real-life materials in teaching lang. & cult.), and 23 (assess successful interaction with classmates). Items 1 (use computer-based technologies), 7 (not use English), 10 (not grade production for accuracy), 11 (have students complete specific tasks rather than grammar), 13 (address errors with immediate explanation), 20 (teach grammar with examples before rules), and 22 (not simplify speech) all resulted in mild agreement with average scores between 2.5 and 3. Teachers disagreed with Items 6 (assess listening and reading via English), 15 (not use small group or pair work), 16 (use activities to practice grammar points rather than information exchange), and 17 (ask students to speak FL only when ready) with the strongest disagreement being expressed on Item 15, which received the lowest score of 1.63. This negatively worded item states that teachers should not use small group or pair work in the foreign language classroom.

Similar to Table 6, Table 7 presents teacher means on each item of the Effective Teacher Questionnaire in rank order from highest to lowest agreement with the question number in parentheses following the item. Items 1 and 20 resulted in the same mean and therefore were both given the same ranking, 16, leading to 23 total rankings.
Table 7. Teacher Means on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Item</th>
<th>Teachers’ Means</th>
<th>“Effective foreign language teachers should . . .”</th>
<th>Teachers’ Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Effective foreign language teachers should . . .&quot; Rank of Item-Item</td>
<td>Teachers’ Means (n = 49)&lt;sup&gt;1&lt;/sup&gt; 1=SD; 4=SA</td>
<td>“Effective foreign language teachers should . . .” Rank of Item-Item</td>
<td>Teachers’ Means (n = 49)&lt;sup&gt;1&lt;/sup&gt; 1=SD; 4=SA</td>
</tr>
<tr>
<td>1-engage students in information gap activities (Q24)</td>
<td>3.51</td>
<td>13-not correct immediately (Q5)</td>
<td>3.02</td>
</tr>
<tr>
<td>2-be as knowledgeable about culture as language (Q9)</td>
<td>3.49</td>
<td>14-have students complete specific tasks rather than grammar (Q11) (n = 48)&lt;sup&gt;1&lt;/sup&gt; 3.00</td>
<td></td>
</tr>
<tr>
<td>3-assess group tasks (Q2)</td>
<td>3.45</td>
<td>15-not grade production for accuracy (Q10)</td>
<td>2.98</td>
</tr>
<tr>
<td>4-have students respond to physical commands (TPR) (Q12)</td>
<td>3.35</td>
<td>16-use computer-based technologies (Q1)</td>
<td>2.94</td>
</tr>
<tr>
<td>5-use real-life materials in teaching lang. &amp; cult. (Q21) (n = 48)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>3.29</td>
<td>16-teach grammar with examples before rules (Q20)</td>
<td>2.94</td>
</tr>
<tr>
<td>6-only correct oral errors indirectly (Q8) (n = 48)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>3.25</td>
<td>17-address errors with immediate explanation (Q13) (n = 48)&lt;sup&gt;1&lt;/sup&gt; 2.71</td>
<td></td>
</tr>
<tr>
<td>7-devote time to culture (Q3)</td>
<td>3.24</td>
<td>18-not simplify speech (Q22) (n = 48)&lt;sup&gt;1&lt;/sup&gt; 2.67</td>
<td></td>
</tr>
<tr>
<td>8-not present grammar without real-world context (Q18) (n = 48)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>3.17</td>
<td>19-not use English (Q7)</td>
<td></td>
</tr>
<tr>
<td>9-assess successful interaction with classmates (Q23)</td>
<td>3.16</td>
<td>20-assess listening and reading via English (Q6) (n = 48)&lt;sup&gt;1&lt;/sup&gt; 2.42</td>
<td></td>
</tr>
<tr>
<td>10-have students use language outside of class (Q4) (n = 47)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>3.15</td>
<td>21-use activities to practice grammar points rather than information exchange (Q16)</td>
<td>2.08</td>
</tr>
<tr>
<td>11-require students</td>
<td>3.14</td>
<td>22-ask students to</td>
<td>2.04</td>
</tr>
</tbody>
</table>
RESEARCH QUESTION 3

Based on a similar 24-item, Likert-scale questionnaire (Effective Teacher Questionnaire), how do students’ and teachers’ perceptions of effective foreign language teachers coincide or differ?

As compared to students’ opinions, teachers’ opinions of the issues on the questionnaire were more polarized, as their responses ranged from 1.63 to 3.51 on a 4-point scale with 1 being the minimum and 4 being the maximum. Students’ responses did not demonstrate such large variation as their scores ranged from 2.03 to 3.25. Since t-tests were run on every item, it was necessary to use a Bonferroni adjustment wherein the
traditional alpha level of .05 was divided by 24, giving an adjusted alpha level of .0021. After this adjustment was made, the differences between teachers and students were statistically significant for exactly half (12) of the items, specifically for the following items: 1 (use computer-based technologies), 2 (assess group tasks), 3 (devote time to culture), 4 (have students use language outside of class), 5 (not correct immediately), 9 (be as knowledgeable about culture as language), 11 (have students complete specific tasks rather than grammar), 13 (address errors with immediate explanation), 14 (require students to speak FL first day of class), 15 (not use small group or pair work), 16 (use activities to practice grammar points rather than information exchange), and 24 (engage students in information gap activities). In spite of apparently small differences between students and teachers on the 4-point scale, the large number of student participants led to statistical power and, thus, many significant differences.

Another approach to analyzing the resulting differences, which may prove more simple to interpret, would be to identify those items that demonstrated the largest, and smallest, discrepancies in raw scores between students and teachers on the 4-point scale. For example, although differences much smaller than .50 resulted in statistical difference, practically speaking, a raw score difference of .50 on a 4-point scale would appear to represent a more extreme difference of opinion, especially given the large number of participants. Items 5 (not correct immediately), 11 (have students complete specific tasks rather than grammar), 14 (require students to speak FL first day of class), 15 (not use small group or pair work), and 16 (use activities to practice grammar points rather than information exchange) each resulted in a difference of .49 or higher. Teachers seem to be
less comfortable with grammar instruction and more amenable to communicative language teaching strategies than students. Teachers were in agreement by more than half a point (> 0.5) that foreign language teachers should require their students to begin speaking the foreign language the first day of class (Item 14) and that engaging in communicative tasks rather than grammar was desirable (Item 11). When asked about whether teachers should focus on specific grammar points rather than on the exchange of information (Item 16), students expressed their agreement while teachers expressed their disagreement resulting in a .64 difference. The largest discrepancy (.90) was found on Item 5, which states “Effective foreign language teachers should not correct students immediately after they make a mistake in speaking.” Students disagreed with this statement while teachers agreed that they should not immediately intervene when students make oral mistakes.

Of the 12 items that were not statistically significant, several questions resulted in minimal differences, i.e., mean differences of less than .15. These six items included the following questions with 17 resulting in a raw teacher-student difference of only .01: 6 (assess listening and reading via English), 17 (ask students to speak FL only when ready), 18 (not present grammar without real-world context), 19 (speak with native-like control of language), 20 (teach grammar with examples before rules), and 22 (not simplify speech). Item 17, with the smallest difference, states that teachers should only ask students to begin speaking the language when they feel they are ready to—a statement both teachers and students disagreed with nearly equally. Item 6 regarding the use of English in assessing reading and listening elicited a neutral response from both groups, as
did Item 22 concerning the need for teachers to not simplify their speech in the classroom. Items 18 and 20 regarding grammar in real-world contexts and inductive approaches to grammar produced moderate agreement with very minimal differences between each group’s responses. Both teachers and students expressed moderate agreement that foreign language teachers should demonstrate native-like control of the language (Item 19) and differed only marginally in their level of agreement—a difference of .13.

Following Feldman’s (1989) practice of analyzing both relative (correlation analyses) and absolute (inertial tests of significance for means differences) similarity between students’ and teachers’ responses, a Pearson correlation coefficient was calculated where the 49 teachers’ responses to each item were correlated with students’ responses on each respective item. A correlation coefficient was generated for all 24 items and resulted in very low correlations per item as manifest in Table 8. These low correlations are not surprising, given that the Effective Teacher Questionnaire elicits students’ and teachers’ perceptions of what they think should be happening, in general, in FL classes. Thus, there would be no reason to believe that their responses would correlate closely with their teachers’ as students and teachers were not asked to reflect on their particular class or on any specific teacher. Therefore, although high correlations could have resulted, the fact that most correlations approach zero is understandable. Table 8 also includes the $p$ value of statistical significance for each correlation indicating whether the correlation is significantly different from zero as well as an indication of how each item’s correlation ranks as compared to the others. Only Item 2 regarding the assessing of
group tasks resulted in significant correlations when all teachers’ responses are compared to all students’ responses. Table 9 presents teachers’ and students’ overall means side-by-side and includes the *p* value which resulted from a t-test between all teachers’ and all students’ responses on each individual item. Additionally, Table 9 includes a column containing the differences between students’ and teachers’ mean responses (Ss-Ts=Mean Difference) on each item and ranks the items in descending order from the largest to the smallest mean difference using absolute values. The negative sign has been retained in order to indicate direction of difference where negative values represent greater teacher than student agreement on the item. Once again, more than one item resulted in similar mean difference values causing two items to be ranked 8 (Items 4 and 24) and two separate items to be ranked 21 (Items 6 and 20).

Table 8. Correlations of Teachers’ and Students’ Responses by Item on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Item Number “Effective foreign language teachers should . . .”</th>
<th>Rank of Item-Item</th>
<th>Correlation Coefficient (r)</th>
<th><em>p</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-assess group tasks (Q2)</td>
<td>.285</td>
<td>.047*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-not simplify speech (Q22)</td>
<td>.261</td>
<td>.074</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-require students to speak FL first day of class (Q14)</td>
<td>.246</td>
<td>.088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-have students use language outside of class (Q4)</td>
<td>-.218</td>
<td>.142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-not use English (Q7)</td>
<td>.210</td>
<td>.148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-devote time to culture (Q3)</td>
<td>-.183</td>
<td>.209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-engage students in information gap activities (Q24)</td>
<td>.175</td>
<td>.229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-assess successful interaction with classmates (Q23)</td>
<td>.169</td>
<td>.247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-address errors with immediate explanation (Q13)</td>
<td>.164</td>
<td>.265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-use computer-based technologies (Q1)</td>
<td>.147</td>
<td>.313</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-be as knowledgeable about culture as language (Q9)</td>
<td>.142</td>
<td>.329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-not grade production for accuracy (Q10)</td>
<td>-.138</td>
<td>.346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-speak with native-like control of language (Q19)</td>
<td>.114</td>
<td>.437</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9. Overall Comparison of Student and Teacher Means by Item on Effective Teacher Questionnaires including p Values from t-tests.

<table>
<thead>
<tr>
<th>“Effective foreign language teachers should . . .” Rank of Item-Item</th>
<th>Teachers’ Means 1=SD, 4=SA n = 49</th>
<th>Students’ Means 1=SD, 4=SA n = 1409</th>
<th>Mean Difference (Ss-Ts)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-not correct immediately (Q5)</td>
<td>3.02</td>
<td>2.12</td>
<td>-0.9</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>2-use activities to practice grammar points rather than information (Q16) exchange</td>
<td>2.08</td>
<td>2.72</td>
<td>0.64</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>3-require students to speak FL first day of class (Q14)</td>
<td>3.14</td>
<td>2.55</td>
<td>-0.59</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>4-have students complete specific tasks rather than grammar (Q11) (n = 48)</td>
<td>3</td>
<td>2.46</td>
<td>-0.54</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>5-devote time to culture (Q3)</td>
<td>3.24</td>
<td>2.74</td>
<td>-0.5</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>6-not use small group or pair work (Q15) (n = 48)</td>
<td>1.63</td>
<td>2.12</td>
<td>0.49</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>7-assess group tasks (Q2)</td>
<td>3.45</td>
<td>2.97</td>
<td>-0.48</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>8-have students use language outside of class (n = 47)</td>
<td>3.15</td>
<td>2.69</td>
<td>-0.46</td>
<td>0.0001*</td>
</tr>
</tbody>
</table>

* significant at an alpha level of .05.
<table>
<thead>
<tr>
<th>(Q4)</th>
<th>Score</th>
<th>Score</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-engage students in information gap activities (Q24)</td>
<td>3.51</td>
<td>3.05</td>
<td>-0.46</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>9-address errors with immediate explanation (Q13)</td>
<td>(n = 48)</td>
<td>2.71</td>
<td>3.13</td>
<td>0.42</td>
</tr>
<tr>
<td>10-use computer-based technologies (Q1)</td>
<td>2.94</td>
<td>2.53</td>
<td>-0.41</td>
<td>0.0001*</td>
</tr>
<tr>
<td>11-be as knowledgeable about culture as language (Q9)</td>
<td>3.49</td>
<td>3.10</td>
<td>-0.39</td>
<td>0.0002*</td>
</tr>
<tr>
<td>12-not grade production for accuracy (Q10)</td>
<td>2.98</td>
<td>2.66</td>
<td>-0.32</td>
<td>0.02</td>
</tr>
<tr>
<td>13-have students respond to physical commands (TPR) (Q12)</td>
<td>3.35</td>
<td>3.06</td>
<td>-0.29</td>
<td>0.0024</td>
</tr>
<tr>
<td>14-not use English (Q7)</td>
<td>2.63</td>
<td>2.36</td>
<td>-0.27</td>
<td>0.0239</td>
</tr>
<tr>
<td>15-only correct oral errors indirectly (Q8)</td>
<td>(n = 48)</td>
<td>3.25</td>
<td>3.02</td>
<td>-0.23</td>
</tr>
<tr>
<td>16-assess successful interaction with classmates (Q23)</td>
<td>3.16</td>
<td>2.95</td>
<td>-0.21</td>
<td>0.0204</td>
</tr>
<tr>
<td>17-use real-life materials in teaching lang. &amp; cult. (Q21)</td>
<td>(n = 48)</td>
<td>3.29</td>
<td>3.11</td>
<td>-0.18</td>
</tr>
<tr>
<td>18-speak with native-like control of language (Q19)</td>
<td>3.12</td>
<td>3.25</td>
<td>0.13</td>
<td>0.2862</td>
</tr>
<tr>
<td>19-not present grammar without real-world context (Q18)</td>
<td>(n = 48)</td>
<td>3.17</td>
<td>3.09</td>
<td>-0.08</td>
</tr>
<tr>
<td>20-not simplify speech (Q22)</td>
<td>(n = 48)</td>
<td>2.67</td>
<td>2.60</td>
<td>-0.07</td>
</tr>
<tr>
<td>21-assess listening and reading via English (Q6)</td>
<td>(n = 48)</td>
<td>2.42</td>
<td>2.36</td>
<td>-0.06</td>
</tr>
<tr>
<td>21-teach grammar with examples before rules (Q20)</td>
<td>2.94</td>
<td>2.88</td>
<td>-0.06</td>
<td>0.5558</td>
</tr>
<tr>
<td>22-ask students to speak FL only when ready (Q17)</td>
<td>2.04</td>
<td>2.03</td>
<td>-0.01</td>
<td>0.9006</td>
</tr>
</tbody>
</table>

* significant at a Bonferroni adjusted alpha level of .0021 to avoid Type I error.
RESEARCH QUESTION 4

How do individual teachers’ perceptions of effective foreign language teachers coincide with or differ from those of their students’?

The analyses up to this point have focused on students and teachers as large independent groups rather than breaking down the responses by teacher. Research Question 4 deals specifically with the degree of similarity between each teacher’s responses and the responses of that teacher’s students on each item. Rather than display the dozens of pages of statistical output that this analysis produced as a result of comparing the responses of 49 teachers to their over 1400 students on over 20 items, summary tables are used to present two separate groups of statistics for Research Question 4 (Effective Teacher Questionnaire), Research Question 5 (frequency column of Evaluation Questionnaire), and Research Question 6 (evaluation column of Evaluation Questionnaire).
The first group of statistics represents averaged differences between individual teacher’s raw response values and his/her students’ means for each item on each questionnaire’s respective 4-point scale. For Research Question 4, each individual teacher’s difference between his/her response on an item and his/her students’ mean responses were averaged across all teachers by item as displayed in Table 10. A negative value indicates that overall teachers were in greater disagreement than students on an item while a positive value indicates greater teacher than student agreement. Table 10 also indicates how many of the 49 teachers were in greater agreement individually than their students and how many were in greater disagreement than their students on each item.

The second group of statistics to be reported for Research Question 4, 5, and 6 concerns correlation analyses, both overall across all teachers and students by item as well as an overall correlation between each teacher and his/her students across all items on each questionnaire. Table 11 presents the overall correlations between each teacher’s responses with their students’ responses across all 24 items on the Effective Teacher Questionnaire. These two groups of statistics, i.e., descriptive statistics of raw teacher-student differences and Pearson product-moment correlations between teachers’ and students’ responses will orient the present analysis and subsequent analyses for Research Questions 5 and 6.

As explained previously, no inferential statistics, i.e., t-tests, were conducted on any comparisons between individual teachers and their students due to the lack of possible variance (n-1) with only one teacher in a means comparison. An additional
factor that prohibits teacher-student comparisons with inferential statistics pertains to the differing stems used for teachers and students on the Evaluation Questionnaire which may have created a separate dependent measure for each group, ruling out a direct comparison of values. In spite of these limitations on the ability to make statistical inferences regarding the effect of probability and chance on teacher-student differences, an attempt is made in this chapter to highlight several specific aspects of teacher-student comparisons for Research Questions 4, 5, and 6 including the following: 1) those items from the questionnaires where the largest and smallest teacher-student differences resulted as well as those teachers with the largest and smallest differences across all items of a questionnaire, 2) those items where the direction of teacher-student difference on an item for individual teachers, i.e., greater teacher than student agreement or vice-versa, clearly pointed to an established tendency among the participating teacher population, 3) those items with high degrees of positive or negative correlation between teachers’ and students’ response patterns along with those teachers with high degrees of positive or negative correlation with their students’ perspectives across all items of a questionnaire.

As displayed in Table 10, several items resulted in an averaged mean difference between individual teachers and their students of .50 or greater on a 4-point scale where a negative value indicates greater teacher than student disagreement with the item. Items 3 (devote time to culture), 5 (not correct immediately), 11 (have students complete specific tasks rather than grammar), and 14 (require students to speak FL first day of class) all resulted in an overwhelming majority of teachers agreeing with the item more than their students. The reverse was true for Items 15 (not use small group or pair work) and 16
(use activities to practice grammar points rather than information exchange) where a large majority of teachers were in greater disagreement than their students on the item. On Item 15, 41 of the total 49 teachers who responded expressed greater disagreement than their students that effective foreign language teachers should not use small or pair work in the FL classroom. Item 5 resulted in the largest mean difference at nearly .9 on a 4-point Likert scale with 39 teachers whose responses demonstrated greater agreement than their students that effective foreign language teachers should not correct their students immediately.

The following items resulted in teacher-student differences of less than .10 and appear to represent areas of minimal difference between teachers and students, but not necessarily agreement with the item: 6 (assess listening and reading via English), 17 (ask students to speak FL only when ready), 18 (not present grammar without real-world context), 20 (teach grammar with examples before rules), and 22 (not simplify speech). Due to the minimal disparity between the averaged differences between individual teachers and students, none of these items demonstrated a clear trend of direction of teacher-student agreement or disagreement as might be expected. That is to say that these items were almost evenly split between those teachers with slightly greater agreement than students and those with slightly greater disagreement than students. Item 20 resulted in the greatest disparity with 29 teachers agreeing more strongly than their students and 20 teachers disagreeing more strongly than their students—however minimal that difference was in either direction.
Following the protocol adopted by other tables, Table 10 ranks the items in descending order from the largest to the smallest averaged difference using absolute values. Similar to Table 9, negative signs are included in the table to indicate direction of difference. At the individual teacher level, a negative difference represents greater teacher disagreement than student disagreement, or greater student than teacher agreement, with the respective item.

Table 10. Summary of Averaged Differences between Individual Teachers’ Raw Scores and their Students’ Mean Responses and Direction of Difference by Item on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>“Effective foreign language teachers should . . .” Rank of Item-Item</th>
<th>Averaged Difference between Individual Teachers’ Raw Scores and their Students’ Mean Scores (Ss-Ts)</th>
<th>Number of Teachers with Greater Agreement than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers with Greater Disagreement than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers Responding to Item and % of Total Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-not correct immediately (Q5)</td>
<td>.897</td>
<td>39 (80%)</td>
<td>10 (20%)</td>
<td>49 (100%)</td>
</tr>
<tr>
<td>2-use activities to practice grammar points rather than information exchange (Q16)</td>
<td>-.634</td>
<td>12 (24%)</td>
<td>37 (76%)</td>
<td>49 (100%)</td>
</tr>
<tr>
<td>3-require students to speak FL first day of class (Q14)</td>
<td>.623</td>
<td>38 (78%)</td>
<td>11 (22%)</td>
<td>49 (100%)</td>
</tr>
<tr>
<td>4-have students complete specific tasks rather than grammar (Q11)</td>
<td>.546</td>
<td>36 (75%)</td>
<td>12 (25%)</td>
<td>48 (98%)</td>
</tr>
<tr>
<td>5-devote time to culture (Q3)</td>
<td>.528</td>
<td>39 (80%)</td>
<td>10 (20%)</td>
<td>49 (100%)</td>
</tr>
<tr>
<td></td>
<td>Method/Task</td>
<td>Rho</td>
<td>Agree (60%)</td>
<td>Disagree (40%)</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------</td>
<td>------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>6</td>
<td>not use small group or pair work (Q15)</td>
<td>-.510</td>
<td>7 (15%)</td>
<td>41 (85%)</td>
</tr>
<tr>
<td>7</td>
<td>assess group tasks (Q2)</td>
<td>.492</td>
<td>37 (76%)</td>
<td>12 (24%)</td>
</tr>
<tr>
<td>8</td>
<td>engage students in information gap activities (Q24)</td>
<td>.476</td>
<td>34 (69%)</td>
<td>15 (31%)</td>
</tr>
<tr>
<td>9</td>
<td>have students use language outside of class (Q4)</td>
<td>.465</td>
<td>35 (74%)</td>
<td>12 (26%)</td>
</tr>
<tr>
<td>10</td>
<td>address errors with immediate explanation (Q13)</td>
<td>-.427</td>
<td>7 (15%)</td>
<td>41 (85%)</td>
</tr>
<tr>
<td>11</td>
<td>use computer-based technologies (Q1)</td>
<td>.422</td>
<td>36 (73%)</td>
<td>13 (27%)</td>
</tr>
<tr>
<td>12</td>
<td>be as knowledgeable about culture as language (Q9)</td>
<td>.420</td>
<td>33 (67%)</td>
<td>16 (33%)</td>
</tr>
<tr>
<td>13</td>
<td>not use English (Q7)</td>
<td>.328</td>
<td>29 (59%)</td>
<td>20 (41%)</td>
</tr>
<tr>
<td>14</td>
<td>not grade production for accuracy (Q10)</td>
<td>.293</td>
<td>34 (69%)</td>
<td>15 (31%)</td>
</tr>
<tr>
<td>15</td>
<td>have students respond to physical commands (TPR) (Q12)</td>
<td>.291</td>
<td>27 (55%)</td>
<td>22 (45%)</td>
</tr>
<tr>
<td>16</td>
<td>assess successful interaction with classmates (Q23)</td>
<td>.237</td>
<td>33 (67%)</td>
<td>16 (33%)</td>
</tr>
<tr>
<td>17</td>
<td>only correct oral errors indirectly (Q8)</td>
<td>.233</td>
<td>30 (63%)</td>
<td>18 (37%)</td>
</tr>
<tr>
<td>18</td>
<td>use real-life materials in teaching lang. &amp;</td>
<td>.189</td>
<td>26 (54%)</td>
<td>22 (46%)</td>
</tr>
<tr>
<td>Item Description</td>
<td>Correlation</td>
<td>Teacher Mean (%)</td>
<td>Student Mean (%)</td>
<td>Average Mean (%)</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>19-speak with native-like control of language (Q19)</td>
<td>-.120</td>
<td>20 (41 %)</td>
<td>29 (59 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>20-not simplify speech (Q22)</td>
<td>.097</td>
<td>24 (50 %)</td>
<td>24 (50 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>21-assess listening and reading via English (Q6)</td>
<td>.081</td>
<td>22 (46 %)</td>
<td>26 (54 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>22-not present grammar without real-world context (Q18)</td>
<td>.069</td>
<td>24 (50 %)</td>
<td>24 (50 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>23-teach grammar with examples before rules (Q20)</td>
<td>.057</td>
<td>29 (59 %)</td>
<td>20 (41 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>24-ask students to speak FL only when ready (Q17)</td>
<td>.023</td>
<td>24 (49 %)</td>
<td>25 (51 %)</td>
<td>49 (100 %)</td>
</tr>
</tbody>
</table>

Tables 11 and 12 present a teacher-by-teacher comparison by displaying, first, each teacher’s averaged teacher-student difference across all 24 items on the Effective Teacher Questionnaire (Table 11), and, second, each teacher’s individual correlation with their students’ responses across all 24 items on the same questionnaire (Table 12). In other words, each teacher’s value on each item is compared to their students’ mean value on the same item to create a difference value per item. The average of these differences across all 24 items is reported in Table 11 for each teacher. Responses from Teachers 16, 20, 43, 45, and 49 across all items on the Effective Teacher questionnaire differed from their students, on average, by .50 or more. While specific items surely demonstrated differences much larger and smaller between individual teachers and their students, this averaged difference provides an idea of which teachers hold different perspectives from
their students on the target teaching behaviors included on the Effective Teacher Questionnaire. In contrast with these five teachers, Teachers 2, 9, 21, 25, and 42 all had averaged teacher-student differences of .05 or lower. Again, negative values represent the direction of difference, more concretely, that teachers were in greater disagreement than their students overall. An interesting finding is that only 7 of the 49 teachers averaged greater disagreement than their students on the 24 items on the Effective Teacher Questionnaire as manifested by the negative averages. Table 11 includes a column displaying the rank for each teacher’s averaged difference from their students in descending order using absolute values.
Table 11. Averaged Teacher-Student Difference across 24 Items on Effective Teacher Questionnaire by Teacher.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across 24 Items on Effective Teacher Questionnaire (4-point scale)</th>
<th>Rank</th>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across 24 Items on Effective Teacher Questionnaire (4-point scale)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.337</td>
<td>14</td>
<td>26</td>
<td>.288</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>-.044</td>
<td>43</td>
<td>27</td>
<td>.306</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>.184</td>
<td>26</td>
<td>28</td>
<td>.141</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>.109</td>
<td>34</td>
<td>29</td>
<td>.155</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>.192</td>
<td>25</td>
<td>30</td>
<td>-.056</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>-.155</td>
<td>28</td>
<td>31</td>
<td>.014</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>.219</td>
<td>21</td>
<td>32</td>
<td>-.073</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>.288</td>
<td>18</td>
<td>33</td>
<td>.088</td>
<td>35</td>
</tr>
<tr>
<td>9</td>
<td>-.015</td>
<td>45</td>
<td>34</td>
<td>.070</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>.150</td>
<td>29</td>
<td>35</td>
<td>.126</td>
<td>33</td>
</tr>
<tr>
<td>11</td>
<td>.357</td>
<td>10</td>
<td>36</td>
<td>.074</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>.197</td>
<td>24</td>
<td>37</td>
<td>.353</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>.130</td>
<td>32</td>
<td>38</td>
<td>.215</td>
<td>22</td>
</tr>
<tr>
<td>14</td>
<td>.496</td>
<td>6</td>
<td>39</td>
<td>.443</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>.220</td>
<td>20</td>
<td>40</td>
<td>.356</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>.644</td>
<td>1</td>
<td>41</td>
<td>.206</td>
<td>23</td>
</tr>
<tr>
<td>17</td>
<td>.471</td>
<td>7</td>
<td>42</td>
<td>-.045</td>
<td>42</td>
</tr>
<tr>
<td>18</td>
<td>.184</td>
<td>26</td>
<td>43</td>
<td>.530</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>.182</td>
<td>27</td>
<td>44</td>
<td>.134</td>
<td>31</td>
</tr>
<tr>
<td>20</td>
<td>.626</td>
<td>2</td>
<td>45</td>
<td>.502</td>
<td>5</td>
</tr>
<tr>
<td>21</td>
<td>.050</td>
<td>41</td>
<td>46</td>
<td>.448</td>
<td>8</td>
</tr>
<tr>
<td>22</td>
<td>-.315</td>
<td>15</td>
<td>47</td>
<td>.342</td>
<td>13</td>
</tr>
<tr>
<td>23</td>
<td>.283</td>
<td>19</td>
<td>48</td>
<td>.087</td>
<td>36</td>
</tr>
<tr>
<td>24</td>
<td>.310</td>
<td>16</td>
<td>49</td>
<td>.566</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>.021</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12 reports correlations between teachers’ responses on all 24 items on the Effective Teacher Questionnaire and their respective students’ responses. Correlations do
not reflect raw differences in response values, rather the degree to which teachers’ responses parallel, or mirror their students’. For this reason many teachers who may have large differences in raw values from their students may actually have high correlations assuming that the patterns of variability across items between a given teacher and students parallel each other. Table 12 presents the Pearson product-moment correlations for each teacher on the Effective Teacher Questionnaire and whether the correlations are significantly different from zero as well the rankings of each teacher’s correlation value. It should be pointed out that even though a correlation is significantly different from zero it may not be very meaningful in the pedagogical sense. That is to say that although a correlation under .50 may be significantly different from zero from a statistical standpoint it may not truly reflect a close relationship between students’ and teachers’ response patterns concretely.

Seventeen of the 49 participating teachers produced significant correlations with their students on the Effective Teacher Questionnaire. Of these 17, seven teachers (Teachers 11, 17, 25, 31, 33, 35, and 47) had responses across all 24 items that correlated at a rate of .50 or higher with Teacher 35’s responses producing a correlation of .733. In order to avoid mention of potentially meaningless correlations from a practical, pedagogical standpoint, only those teachers with correlations of .50 or higher were highlighted. Only one teacher, Teacher 3, had a negative correlation—a value which proved to not be significantly different from zero.
Table 12. Pearson Product-moment Correlation for Individual Teachers and their Students on Effective Teacher Questionnaire.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Effective Teacher Questionnaire</th>
<th>Rank</th>
<th>p value</th>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Effective Teacher Questionnaire</th>
<th>Rank</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.357</td>
<td>25</td>
<td>.087</td>
<td>26</td>
<td>.421</td>
<td>15</td>
<td>.040</td>
</tr>
<tr>
<td>2</td>
<td>.276</td>
<td>32</td>
<td>.192</td>
<td>27</td>
<td>.077</td>
<td>44</td>
<td>.721</td>
</tr>
<tr>
<td>3</td>
<td>-.006</td>
<td>47</td>
<td>.979</td>
<td>28</td>
<td>.163</td>
<td>40</td>
<td>.447</td>
</tr>
<tr>
<td>4</td>
<td>.493</td>
<td>9</td>
<td>.014*</td>
<td>29</td>
<td>.297</td>
<td>30</td>
<td>.158</td>
</tr>
<tr>
<td>5</td>
<td>.463</td>
<td>11</td>
<td>.023*</td>
<td>30</td>
<td>.285</td>
<td>31</td>
<td>.176</td>
</tr>
<tr>
<td>6</td>
<td>.121</td>
<td>43</td>
<td>.573</td>
<td>31</td>
<td>.534</td>
<td>7</td>
<td>.007*</td>
</tr>
<tr>
<td>7</td>
<td>.395</td>
<td>21</td>
<td>.056</td>
<td>32</td>
<td>.417</td>
<td>16</td>
<td>.042*</td>
</tr>
<tr>
<td>8</td>
<td>.379</td>
<td>22</td>
<td>.068</td>
<td>33</td>
<td>.622</td>
<td>4</td>
<td>.001*</td>
</tr>
<tr>
<td>9</td>
<td>.364</td>
<td>24</td>
<td>.081</td>
<td>34</td>
<td>.398</td>
<td>20</td>
<td>.054</td>
</tr>
<tr>
<td>10</td>
<td>.304</td>
<td>28</td>
<td>.149</td>
<td>35</td>
<td>.733</td>
<td>1</td>
<td>.000*</td>
</tr>
<tr>
<td>11</td>
<td>.642</td>
<td>3</td>
<td>.001*</td>
<td>36</td>
<td>.049</td>
<td>45</td>
<td>.825</td>
</tr>
<tr>
<td>12</td>
<td>.186</td>
<td>39</td>
<td>.408</td>
<td>37</td>
<td>.365</td>
<td>23</td>
<td>.079</td>
</tr>
<tr>
<td>13</td>
<td>.186</td>
<td>39</td>
<td>.396</td>
<td>38</td>
<td>.456</td>
<td>12</td>
<td>.025*</td>
</tr>
<tr>
<td>14</td>
<td>.409</td>
<td>18</td>
<td>.047*</td>
<td>39</td>
<td>.468</td>
<td>10</td>
<td>.021*</td>
</tr>
<tr>
<td>15</td>
<td>.410</td>
<td>17</td>
<td>.047*</td>
<td>40</td>
<td>.408</td>
<td>19</td>
<td>.048</td>
</tr>
<tr>
<td>16</td>
<td>.131</td>
<td>42</td>
<td>.540</td>
<td>41</td>
<td>.197</td>
<td>38</td>
<td>.356</td>
</tr>
<tr>
<td>17</td>
<td>.545</td>
<td>6</td>
<td>.006*</td>
<td>42</td>
<td>.348</td>
<td>26</td>
<td>.095</td>
</tr>
<tr>
<td>18</td>
<td>.437</td>
<td>14</td>
<td>.033*</td>
<td>43</td>
<td>.285</td>
<td>31</td>
<td>.177</td>
</tr>
<tr>
<td>19</td>
<td>.440</td>
<td>13</td>
<td>.032*</td>
<td>44</td>
<td>.275</td>
<td>33</td>
<td>.205</td>
</tr>
<tr>
<td>20</td>
<td>.497</td>
<td>8</td>
<td>.013*</td>
<td>45</td>
<td>.327</td>
<td>27</td>
<td>.119</td>
</tr>
<tr>
<td>21</td>
<td>.015</td>
<td>46</td>
<td>.944</td>
<td>46</td>
<td>.233</td>
<td>36</td>
<td>.284</td>
</tr>
<tr>
<td>22</td>
<td>.214</td>
<td>37</td>
<td>.315</td>
<td>47</td>
<td>.704</td>
<td>2</td>
<td>.000*</td>
</tr>
<tr>
<td>23</td>
<td>.300</td>
<td>29</td>
<td>.154</td>
<td>48</td>
<td>.240</td>
<td>35</td>
<td>.295</td>
</tr>
<tr>
<td>24</td>
<td>.134</td>
<td>41</td>
<td>.532</td>
<td>49</td>
<td>.270</td>
<td>34</td>
<td>.203</td>
</tr>
<tr>
<td>25</td>
<td>.577</td>
<td>5</td>
<td>.004*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significant at an alpha level of .05.
RESEARCH QUESTION 5

How do students’ and teachers’ perceptions of the frequency of occurrence of certain teaching behaviors in their classrooms coincide or differ, both overall and in a teacher-by-teacher comparison on a 21-item, Likert-scale questionnaire (Evaluation Questionnaire)?

Another crucial objective of this study is to compare and contrast students’ and teachers’ perceptions of what is actually happening in the classroom. Research Question 5 asked students and teachers to reflect on how often teachers performed certain behaviors in the classroom. The items on these questionnaires (see Appendix F and G) included two columns as described in Chapter 3, the first asking how often a teacher engaged in a certain practice and the second requesting an evaluation of the teacher’s performance of that specific teaching practice. Do students perceive certain behaviors to be taking place at the same general rate as their teachers? Clearly, the nature of the adverbs used on the questionnaire such as frequently, sometimes, and rarely may be interpreted differently for each individual. Nevertheless, the goal of the research study is not to objectively quantify the exact frequency of occurrence of the target teaching behaviors, but rather to compare subjective perceptions. For example, does Teacher ‘X’ feel that he or she does something “Frequently” while his or her students feel that it is done “Rarely” or “Never”? Does the reverse situation present itself, where teachers feel they do something rarely or never and students perceive it to be executed frequently?

Due to the distinctive nature of students’ and teachers’ versions of the Evaluation Questionnaire, students’ and teachers’ perceptions of the frequency of certain teaching
behaviors were compared using descriptive statistics and correlation analyses as described previously in Research Question 4. Table 13 presents a comparison between all teachers’ and all students’ responses when averaged across items. Also included is the correlation between students’ and teachers’ responses on the frequency column of the Evaluation Questionnaire with the accompanying p value indicating whether the correlation is significantly different from zero. The final column of the table contains the mean difference between all students and all teachers as well as the rank of the mean differences in descending order.

Thirteen items resulted in significant correlations and include the following items: 1 (use/s computer-based technologies), 2 (devote/s time for culture), 3 (has/have students use language outside of class), 4 (correct/s oral mistakes directly), 5 (make/s use of English), 7 (grade/s for grammatical accuracy), 11 (encourage/s students to speak the FL in class), 12 (use/s small-group work), 13 (use/s activities to exchange information rather than grammar), 14 (present/s grammar in real-world context), 16 (teach/es grammar with rules before examples), 17 (use/s real-life materials in the classroom), and 19 (engage/s students in information gap activities). The highest correlation resulted on Item 3 (has/have students use language outside of class) and produced a correlation of r = .642 while only one item resulted in a negative correlation that was not significantly different from zero—Item 21 (use/s FL in the classroom).

Two comments regarding tendencies in teacher-student correlations and teacher-student differences between questionnaires merit mention here. First, the amount of statistical power produced by a large number of subjects as reflected in this study
produces correlations that may be relatively low but significantly different from zero, e.g., $r = .292$, $p = .042$ on Item 7. Second, the number of significant correlations increased dramatically from the Effective Teacher Questionnaire to the Evaluation Questionnaire while the amount of raw teacher-student differences decreased for reasons mentioned earlier—students and teachers are not reflecting on an abstract ideal on the Evaluation Questionnaire, rather their teachers’ concrete performance. Therefore, it seems logical that these values would shift in the direction they have since teachers and students have been asked to reflect on something they both have witnessed rather than the abstraction of “effective foreign language teachers”.

Table 13. Overall Comparison and Correlations of Students’ and Teachers’ Responses on Frequency Column of Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>“In this class my foreign language teacher. . .”/ “As a foreign language teacher, I . . .”</th>
<th>Correlation Coefficient ($r$)</th>
<th>$p$ value</th>
<th>Teacher Means $1 = \text{Never}; 4 = \text{Frequently}$</th>
<th>Student Means $1 = \text{Never}; 4 = \text{Frequently}$</th>
<th>Mean Difference (Ss-Ts); Rank of Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-correct(s) oral mistakes directly (Q4)</td>
<td>.555</td>
<td>.000*</td>
<td>2.73</td>
<td>3.21</td>
<td>0.48</td>
</tr>
<tr>
<td>2-has/have students respond to physical commands—TPR (Q9)</td>
<td>.106</td>
<td>.467</td>
<td>3.05</td>
<td>2.66</td>
<td>-0.39</td>
</tr>
<tr>
<td>3-correct(s) oral mistakes indirectly (Q8)</td>
<td>.177</td>
<td>.223</td>
<td>3.69</td>
<td>3.34</td>
<td>-0.35</td>
</tr>
<tr>
<td>4-teach(es) grammar with rules before examples (Q16)</td>
<td>.409</td>
<td>.004*</td>
<td>(n = 47)$^1$</td>
<td>2.89</td>
<td>3.22</td>
</tr>
<tr>
<td>5-has/have students complete specific tasks with real life purpose</td>
<td>.216</td>
<td>.137</td>
<td>2.93</td>
<td>2.62</td>
<td>-0.31</td>
</tr>
<tr>
<td>(Q6)</td>
<td></td>
<td></td>
<td>(n = 48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---</td>
<td>---</td>
<td>----------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>6-has/have students use language outside of class (Q3)</td>
<td>.642</td>
<td>.000*</td>
<td>2.02</td>
<td>2.29</td>
<td>0.27</td>
</tr>
<tr>
<td>7-alter(s) speech for understanding (Q18)</td>
<td>.177</td>
<td>.235</td>
<td>(n = 47)</td>
<td>3.45</td>
<td>3.19</td>
</tr>
<tr>
<td>8-use(s) real-life materials in the classroom (Q17)</td>
<td>.405</td>
<td>.004*</td>
<td>2.95</td>
<td>2.73</td>
<td>-0.22</td>
</tr>
<tr>
<td>9-assess(es) interaction with classmates (Q20)</td>
<td>.056</td>
<td>.706</td>
<td>(n = 48)</td>
<td>2.60</td>
<td>2.81</td>
</tr>
<tr>
<td>10-devote(s) time for culture (Q2)</td>
<td>.525</td>
<td>.000*</td>
<td>(n = 48)</td>
<td>3.67</td>
<td>3.50</td>
</tr>
<tr>
<td>11-use(s) computer-based technologies (Q1)</td>
<td>.560</td>
<td>.000*</td>
<td>2.03</td>
<td>2.17</td>
<td>0.14</td>
</tr>
<tr>
<td>11-encourage(s) students to speak the FL in class (Q11)</td>
<td>.323</td>
<td>.027*</td>
<td>(n = 47)</td>
<td>3.81</td>
<td>3.67</td>
</tr>
<tr>
<td>12-grade(s) for grammatical accuracy (Q7)</td>
<td>.292</td>
<td>.042*</td>
<td>3.52</td>
<td>3.64</td>
<td>0.12</td>
</tr>
<tr>
<td>13-demonstrate(s) native-like control in FL (Q15)</td>
<td>.225</td>
<td>.125</td>
<td>(n = 48)</td>
<td>3.88</td>
<td>3.77</td>
</tr>
<tr>
<td>14-use(s) activities to exchange information rather than grammar (Q13)</td>
<td>.333</td>
<td>.019*</td>
<td>3.43</td>
<td>3.34</td>
<td>-0.09</td>
</tr>
<tr>
<td>15-present(s) grammar in real-world context (Q14)</td>
<td>.296</td>
<td>.039*</td>
<td>3.26</td>
<td>3.22</td>
<td>-0.04</td>
</tr>
<tr>
<td>15-use(s) FL in the classroom (Q21)</td>
<td>-.074</td>
<td>.615</td>
<td>3.98</td>
<td>3.94</td>
<td>-0.04</td>
</tr>
<tr>
<td>16-address(es) errors by explaining why (Q10)</td>
<td>.097</td>
<td>.508</td>
<td>3.09</td>
<td>3.12</td>
<td>0.03</td>
</tr>
<tr>
<td>16-use(s) small-group work (Q12)</td>
<td>.430</td>
<td>.002*</td>
<td>3.76</td>
<td>3.73</td>
<td>-0.03</td>
</tr>
<tr>
<td>17-make(s) use of English (Q5)</td>
<td>.582</td>
<td>.000*</td>
<td>2.85</td>
<td>2.86</td>
<td>0.01</td>
</tr>
<tr>
<td>18-engage(s) students in information gap</td>
<td>.441</td>
<td>.002*</td>
<td>(n = 48)</td>
<td>3.07</td>
<td>3.07</td>
</tr>
</tbody>
</table>
activities (Q19)

some teachers did not respond to certain items resulting in a lower ‘n’ value.

* significant at an alpha level of .05.

Five of the 21 items on the frequency column of the Evaluation Questionnaire resulted in averaged differences between individual teachers and their students of .30 or greater on a 4-point scale as shown in Table 14. These items included the following: 4 (correct/s oral mistakes directly), 6 (has/have students complete specific tasks with real life purpose), 8 (correct/s oral mistakes indirectly), 9 (has/have students respond to physical commands—TPR), and 16 (teach/es grammar with rules before examples). The largest teacher-student difference resulted on Item 4 relative to how often teachers correct oral mistakes directly with a difference of -.478, meaning that on average, teachers perceived their integration of this strategy as less frequent than did their students. In fact, 39 of the 49 participating teachers felt that they corrected oral mistakes directly less frequently than did their students. Item 16 (teach/es grammar with rules before examples) reflected a similar direction of difference with more teachers perceiving a lower frequency of implementation than did their students. Items 6 (has/have students complete specific tasks with real life purpose), 8 (correct/s oral mistakes indirectly), and 9 (has/have students respond to physical commands—TPR) resulted in more teachers perceiving their integration of these behaviors as occurring more frequently than their students. In contrast to the items with larger differences, Items 10 (address/es errors by explaining why), 12 (use/s small-group work), and 19 (engage/s students in information gap activities) all resulted in teacher-student differences of less than .05 indicating apparent agreement between teachers and students in regard to teachers’ frequency of
implementation of these teaching behaviors. Table 14 organizes the items according to
the rank of the averaged individual differences between teacher’s and their students in
descending order using absolute values. Negative signs indicate that teachers perceived
lower frequency on the item than did their students, on average.

Table 14. Summary of Averaged Differences between Individual Teachers’ Raw Scores
and their Students’ Mean Responses and Direction of Difference by Item on Frequency
Column of Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>“In this class my foreign language teacher. . .”/ “As a foreign language teacher, I . . .” Rank of Item-Item</th>
<th>Averaged Difference between Individual Teachers’ Raw Scores and their Students’ Mean Scores</th>
<th>Number of Teachers with Greater Perceived Frequency than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers with Lower Perceived Frequency than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers Responding to Item and % of Total Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-correct(s) oral mistakes directly (Q4)</td>
<td>-.478</td>
<td>10 (20 %)</td>
<td>39 (80 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>2-has/have students respond to physical commands—TPR (Q9)</td>
<td>.422</td>
<td>32 (65 %)</td>
<td>17 (35 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>3-teach(es) grammar with rules before examples (Q16)</td>
<td>-.377</td>
<td>19 (40 %)</td>
<td>28 (60 %)</td>
<td>47 (96 %)</td>
</tr>
<tr>
<td>4-correct(s) oral mistakes indirectly (Q8)</td>
<td>.357</td>
<td>36 (73 %)</td>
<td>13 (27 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>5-has/have students complete specific tasks with real life purpose (Q6)</td>
<td>.337</td>
<td>31 (63 %)</td>
<td>18 (37 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>6-alter(s) speech for understanding (Q18)</td>
<td>.239</td>
<td>25 (53 %)</td>
<td>22 (47 %)</td>
<td>47 (96 %)</td>
</tr>
<tr>
<td>7-has/have students use language outside of class (Q3)</td>
<td>-.227</td>
<td>18 (38 %)</td>
<td>30 (62 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>8-use(s) real-life materials in the classroom (Q17)</td>
<td>.222</td>
<td>28 (57 %)</td>
<td>21 (%)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>9-assess(es) interaction with classmates (Q20)</td>
<td>-.204</td>
<td>20 (42 %)</td>
<td>28 (58 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>10-use(s) computer-based technologies (Q1)</td>
<td>-.160</td>
<td>16 (33 %)</td>
<td>33 (67 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>11-devote(s) time for culture (Q2)</td>
<td>.137</td>
<td>38 (79 %)</td>
<td>10 (21 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>12-encourage(s) students to speak the FL in class (Q11)</td>
<td>.127</td>
<td>36 (77 %)</td>
<td>11 (23 %)</td>
<td>47 (96 %)</td>
</tr>
<tr>
<td>13-grade(s) for grammatical accuracy (Q7)</td>
<td>-.120</td>
<td>29 (59 %)</td>
<td>20 (41 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>14-demonstrate(s) native-like control in FL (Q15)</td>
<td>.109</td>
<td>40 (83 %)</td>
<td>8 (17 %)</td>
<td>48 (98 %)</td>
</tr>
<tr>
<td>15-present(s) grammar in real-world context (Q14)</td>
<td>.094</td>
<td>25 (51 %)</td>
<td>24 (49 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>16-use(s) activities to exchange information rather than grammar (Q13)</td>
<td>.073</td>
<td>30 (61 %)</td>
<td>19 (39 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>17-make(s) use of English (Q5)</td>
<td>-.058</td>
<td>22 (45 %)</td>
<td>27 (55 %)</td>
<td>49 (100 %)</td>
</tr>
<tr>
<td>18-use(s) FL in the classroom (Q21)</td>
<td>.052</td>
<td>30 (61 %)</td>
<td>19 (39 %)</td>
<td>49 (100 %)</td>
</tr>
</tbody>
</table>
When analyzed from the individual teacher’s perspective, Teachers 17, 26, and 28 all had differences of .40 or greater with Teachers 17 and 28 perceiving they availed themselves of the target teaching techniques with greater frequency, on average, than did their students (positive value). Of these three teachers with teacher-student differences over .40, Teacher 26 was the only one who perceived a lesser frequency than did students when differences were averaged across all 21 items on the frequency column of the Evaluation Questionnaire. The smallest teacher-student differences were manifest in Teachers 12, 19, 39, and 41 whose averaged responses differed from their students’ means by less than .03. Table 15 presents these averaged individual teacher’s differences with students across all 21 items of the frequency column of the Evaluation Questionnaire and includes a column with the rank of each teacher-student difference across all 49 teachers.
Table 15. Averaged Teacher-Student Difference across 21 Items on Frequency Column of Evaluation Questionnaire by Teacher.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across all 21 Items on Frequency Column of Evaluation Questionnaire (4-point scale)</th>
<th>Rank</th>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across all 21 Items on Frequency Column of Evaluation Questionnaire (4-point scale)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.393</td>
<td>4</td>
<td>26</td>
<td>-.524</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-.187</td>
<td>18</td>
<td>27</td>
<td>-.151</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>-.040</td>
<td>42</td>
<td>28</td>
<td>.403</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>.157</td>
<td>23</td>
<td>29</td>
<td>.088</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>-.190</td>
<td>16</td>
<td>30</td>
<td>-.105</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>-.033</td>
<td>43</td>
<td>31</td>
<td>-.120</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>.084</td>
<td>37</td>
<td>32</td>
<td>.369</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>-.265</td>
<td>11</td>
<td>33</td>
<td>-.164</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>.093</td>
<td>34</td>
<td>34</td>
<td>.188</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>.325</td>
<td>8</td>
<td>35</td>
<td>-.137</td>
<td>28</td>
</tr>
<tr>
<td>11</td>
<td>.356</td>
<td>6</td>
<td>36</td>
<td>-.058</td>
<td>39</td>
</tr>
<tr>
<td>12</td>
<td>.009</td>
<td>47</td>
<td>37</td>
<td>-.095</td>
<td>33</td>
</tr>
<tr>
<td>13</td>
<td>-.200</td>
<td>14</td>
<td>38</td>
<td>.123</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>.253</td>
<td>12</td>
<td>39</td>
<td>.013</td>
<td>46</td>
</tr>
<tr>
<td>15</td>
<td>.192</td>
<td>15</td>
<td>40</td>
<td>.147</td>
<td>26</td>
</tr>
<tr>
<td>16</td>
<td>-.175</td>
<td>20</td>
<td>41</td>
<td>-.021</td>
<td>45</td>
</tr>
<tr>
<td>17</td>
<td>.624</td>
<td>1</td>
<td>42</td>
<td>-.139</td>
<td>27</td>
</tr>
<tr>
<td>18</td>
<td>.178</td>
<td>19</td>
<td>43</td>
<td>.060</td>
<td>38</td>
</tr>
<tr>
<td>19</td>
<td>.022</td>
<td>44</td>
<td>44</td>
<td>-.054</td>
<td>40</td>
</tr>
<tr>
<td>20</td>
<td>.151</td>
<td>24</td>
<td>45</td>
<td>.125</td>
<td>29</td>
</tr>
<tr>
<td>21</td>
<td>.210</td>
<td>13</td>
<td>46</td>
<td>.336</td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td>-.054</td>
<td>40</td>
<td>47</td>
<td>.092</td>
<td>35</td>
</tr>
<tr>
<td>23</td>
<td>.148</td>
<td>25</td>
<td>48</td>
<td>-.283</td>
<td>10</td>
</tr>
<tr>
<td>24</td>
<td>-.171</td>
<td>21</td>
<td>49</td>
<td>.287</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>-.052</td>
<td>41</td>
<td>47</td>
<td>.336</td>
<td>7</td>
</tr>
</tbody>
</table>

As Table 16 attests, a mere five of 49 teachers’ responses did not correlate significantly with their students across all 21 items on the frequency column of the
Evaluation Questionnaire. Those teachers whose responses did not correlate significantly were 17, 21, 30, 33, and 45. The overwhelming number of teachers with significant correlations with their students reinforces what has already been stated—teachers and students seem to perceive somewhat similarly how often certain teaching behaviors are taking place. The same conclusion does not hold true for evaluations as evidenced in the following section reviewing the findings for Research Question 6. Table 16 not only includes the Pearson product-moment correlations and the accompanying $p$ values, but also contains a column providing the rank of each teacher’s correlation with their students across all 49 teachers.
Table 16. Pearson Product-moment Correlation for Individual Teachers and their Students on Frequency Column of Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Frequency Column of Evaluation Questionnaire</th>
<th>Rank</th>
<th>p value</th>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Frequency Column of Evaluation Questionnaire</th>
<th>Rank</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.824</td>
<td>3</td>
<td>.000*</td>
<td>26</td>
<td>.709</td>
<td>17</td>
<td>.000*</td>
</tr>
<tr>
<td>2</td>
<td>.600</td>
<td>28</td>
<td>.004*</td>
<td>27</td>
<td>.567</td>
<td>32</td>
<td>.007*</td>
</tr>
<tr>
<td>3</td>
<td>.527</td>
<td>38</td>
<td>.017*</td>
<td>28</td>
<td>.546</td>
<td>35</td>
<td>.010*</td>
</tr>
<tr>
<td>4</td>
<td>.854</td>
<td>2</td>
<td>.000*</td>
<td>29</td>
<td>.757</td>
<td>9</td>
<td>.000*</td>
</tr>
<tr>
<td>5</td>
<td>.783</td>
<td>7</td>
<td>.000*</td>
<td>30</td>
<td>.356</td>
<td>45</td>
<td>.113</td>
</tr>
<tr>
<td>6</td>
<td>.724</td>
<td>16</td>
<td>.000*</td>
<td>31</td>
<td>.670</td>
<td>21</td>
<td>.001*</td>
</tr>
<tr>
<td>7</td>
<td>.538</td>
<td>36</td>
<td>.012*</td>
<td>32</td>
<td>.506</td>
<td>39</td>
<td>.023*</td>
</tr>
<tr>
<td>8</td>
<td>.666</td>
<td>22</td>
<td>.001*</td>
<td>33</td>
<td>.431</td>
<td>44</td>
<td>.051</td>
</tr>
<tr>
<td>9</td>
<td>.565</td>
<td>33</td>
<td>.008*</td>
<td>34</td>
<td>.636</td>
<td>23</td>
<td>.002*</td>
</tr>
<tr>
<td>10</td>
<td>.534</td>
<td>37</td>
<td>.015*</td>
<td>35</td>
<td>.781</td>
<td>8</td>
<td>.000*</td>
</tr>
<tr>
<td>11</td>
<td>.631</td>
<td>24</td>
<td>.002*</td>
<td>36</td>
<td>.746</td>
<td>12</td>
<td>.000*</td>
</tr>
<tr>
<td>12</td>
<td>.736</td>
<td>13</td>
<td>.000*</td>
<td>37</td>
<td>.810</td>
<td>5</td>
<td>.000*</td>
</tr>
<tr>
<td>13</td>
<td>.748</td>
<td>11</td>
<td>.000*</td>
<td>38</td>
<td>.612</td>
<td>26</td>
<td>.003*</td>
</tr>
<tr>
<td>14</td>
<td>.504</td>
<td>40</td>
<td>.020*</td>
<td>39</td>
<td>.700</td>
<td>20</td>
<td>.000*</td>
</tr>
<tr>
<td>15</td>
<td>.731</td>
<td>14</td>
<td>.000*</td>
<td>40</td>
<td>.630</td>
<td>25</td>
<td>.002*</td>
</tr>
<tr>
<td>16</td>
<td>.611</td>
<td>27</td>
<td>.003*</td>
<td>41</td>
<td>.598</td>
<td>29</td>
<td>.004*</td>
</tr>
<tr>
<td>17</td>
<td>.330</td>
<td>46</td>
<td>.196</td>
<td>42</td>
<td>.797</td>
<td>6</td>
<td>.000*</td>
</tr>
<tr>
<td>18</td>
<td>.547</td>
<td>34</td>
<td>.010*</td>
<td>43</td>
<td>.579</td>
<td>30</td>
<td>.006*</td>
</tr>
<tr>
<td>19</td>
<td>.817</td>
<td>4</td>
<td>.000*</td>
<td>44</td>
<td>.727</td>
<td>15</td>
<td>.000*</td>
</tr>
<tr>
<td>20</td>
<td>.749</td>
<td>10</td>
<td>.000*</td>
<td>45</td>
<td>.104</td>
<td>48</td>
<td>.672</td>
</tr>
<tr>
<td>21</td>
<td>.219</td>
<td>47</td>
<td>.340</td>
<td>46</td>
<td>.860</td>
<td>1</td>
<td>.000*</td>
</tr>
<tr>
<td>22</td>
<td>.572</td>
<td>31</td>
<td>.007*</td>
<td>47</td>
<td>.636</td>
<td>23</td>
<td>.002*</td>
</tr>
<tr>
<td>23</td>
<td>.434</td>
<td>43</td>
<td>.049*</td>
<td>48</td>
<td>.705</td>
<td>19</td>
<td>.001*</td>
</tr>
<tr>
<td>24</td>
<td>.708</td>
<td>18</td>
<td>.000*</td>
<td>49</td>
<td>.464</td>
<td>41</td>
<td>.034*</td>
</tr>
<tr>
<td>25</td>
<td>.439</td>
<td>42</td>
<td>.046*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significant at an alpha level of .05.
RESEARCH QUESTION 6

How do teachers’ self-evaluations coincide with or differ from their students’ evaluations of them overall and individually?

This research question compares teachers’ self-evaluations to their students’ evaluations of them overall and individually on the target teaching practices. As explained previously, students and teachers were instructed to reflect on the frequency of occurrence of certain behaviors in the classroom, as well as their evaluations of the concrete execution by their respective teachers on the same questionnaire in two separate columns (Appendices D & E). If they felt that the behavior was “Never” performed then they were instructed to skip the second column and mark “NA” meaning not applicable. Some students misinterpreted the instructions and marked an evaluation after indicating “Never” in the first column asking for frequency. In preparing the data for analysis, all evaluative responses from students who incorrectly providing an evaluation after indicating that they “Never” observed a behavior from a teacher were deleted and not included in the analysis of students’ evaluations. However, their “Never” responses on the frequency column were included in statistical calculations for Research Question 5 which focused on the frequency of certain behaviors. It was apparent that these students were reflecting on the general effectiveness of the behavior rather than their teachers’ concrete performance of the practice, an issue already covered in the Effective Teacher questionnaire administered during the first visit. Teachers were also asked not to evaluate behaviors they felt they “Never” performed, and, therefore, the n values in Table 20 below vary from item to item. For example, only 27 teachers (55%) perceived that they
integrated computer-based technologies in their teaching and only 31 (63%) teachers felt that they required students to use the language outside of class.

Table 17 organizes students’ and teachers’ overall responses similar to previous tables by providing the items in rank order from the largest to the smallest mean difference per item. Although all 21 items are ranked, several items shared the same absolute value of the mean difference with another item and thus received the same rank. Overall, all teachers and all students responses correlated significantly on only two items as demonstrated in Table 17: 9 (has/ have students respond to physical commands—TPR) and 18 (alter/ s speech for understanding). Both of these correlations were positive and, although significantly different from zero, were lower than .50. Table 18 illustrates the average of individual teacher’s raw value differences with their students’ mean per item and is organized by rank order similar to Table 17. Teacher-student differences on the evaluation column were much smaller than on the Effective Teacher Questionnaire, and even slightly more so than the frequency column of the Evaluation Questionnaire.

Table 17. Overall Comparison and Correlations of Students’ and Teachers’ Responses on Evaluation Column on Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>“In this class my foreign language teacher . . .”/ “As a foreign language teacher, I . . .”</th>
<th>Rank of Mean Difference-Item</th>
<th>Correlation Coefficient (r)</th>
<th>p value</th>
<th>Teacher Means 1 = Ineffective; 4 = Very Effective</th>
<th>Student Means 1 = Ineffective; 4 = Very Effective</th>
<th>Mean Difference (Ss-Ts); Rank of Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-correct(s) oral mistakes directly (Q4)</td>
<td>.134</td>
<td>.381</td>
<td>(n = 45)</td>
<td>2.7</td>
<td>3.25</td>
<td>0.55</td>
</tr>
<tr>
<td>2-use(s) computer-based technologies</td>
<td>.308</td>
<td>.118</td>
<td>(n = 27)</td>
<td>3.02</td>
<td>2.75</td>
<td>-0.27</td>
</tr>
<tr>
<td>(Q1)</td>
<td>(Q6)</td>
<td>(Q7)</td>
<td>(Q12)</td>
<td>(Q3)</td>
<td>(Q16)</td>
<td>(Q17)</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>2-has/have students complete specific tasks with real life purpose</td>
<td>.139</td>
<td>.373</td>
<td>(n = 43)</td>
<td>3.28</td>
<td>3.01</td>
<td>-0.27</td>
</tr>
<tr>
<td>2-grade(s) for grammatical accuracy</td>
<td>.243</td>
<td>.093</td>
<td>(n = 49)</td>
<td>3.03</td>
<td>3.30</td>
<td>0.27</td>
</tr>
<tr>
<td>3-uses small-group work (Q12)</td>
<td>.247</td>
<td>.087</td>
<td>(n = 49)</td>
<td>3.5</td>
<td>3.25</td>
<td>-0.25</td>
</tr>
<tr>
<td>4-has/have students use language outside of class (Q3)</td>
<td>-.182</td>
<td>.326</td>
<td>(n = 31)</td>
<td>2.53</td>
<td>2.73</td>
<td>0.2</td>
</tr>
<tr>
<td>5-use(s) real-life materials in the classroom (Q17)</td>
<td>.070</td>
<td>.646</td>
<td>(n = 46)</td>
<td>3.21</td>
<td>3.06</td>
<td>-0.15</td>
</tr>
<tr>
<td>6-correct(s) oral mistakes indirectly</td>
<td>.141</td>
<td>.335</td>
<td>(n = 49)</td>
<td>3.14</td>
<td>3.28</td>
<td>0.14</td>
</tr>
<tr>
<td>6-engage(s) students in information gap activities (Q19)</td>
<td>.147</td>
<td>.340</td>
<td>(n = 44)</td>
<td>3.20</td>
<td>3.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>7-make(s) use of English (Q5)</td>
<td>.240</td>
<td>.101</td>
<td>(n = 47)</td>
<td>3.08</td>
<td>3.21</td>
<td>0.13</td>
</tr>
<tr>
<td>7-encourage(s) students to speak the FL in class (Q11)</td>
<td>.082</td>
<td>.583</td>
<td>(n = 47)</td>
<td>3.15</td>
<td>3.28</td>
<td>0.13</td>
</tr>
<tr>
<td>8-devote(s) time for culture (Q2)</td>
<td>-.011</td>
<td>.943</td>
<td>(n = 48)</td>
<td>3.28</td>
<td>3.17</td>
<td>-0.11</td>
</tr>
<tr>
<td>8-has/have students respond to physical commands—TPR (Q9)</td>
<td>.360</td>
<td>.014*</td>
<td>(n = 46)</td>
<td>3.27</td>
<td>3.16</td>
<td>-0.11</td>
</tr>
<tr>
<td>8-teach(es) grammar with rules before examples (Q16)</td>
<td>.123</td>
<td>.437</td>
<td>(n = 41)</td>
<td>3.07</td>
<td>3.18</td>
<td>0.11</td>
</tr>
<tr>
<td>8-assess(es) interaction with classmates (Q20)</td>
<td>.115</td>
<td>.473</td>
<td>(n = 42)</td>
<td>2.86</td>
<td>2.97</td>
<td>0.11</td>
</tr>
<tr>
<td>9-address(es) errors by explaining why (Q10)</td>
<td>-.077</td>
<td>.601</td>
<td>(n = 48)</td>
<td>3.10</td>
<td>3.19</td>
<td>0.09</td>
</tr>
<tr>
<td>10-present(s) grammar in real-world context (Q14)</td>
<td>-.094</td>
<td>.522</td>
<td>(n = 49)</td>
<td>3.27</td>
<td>3.21</td>
<td>-0.06</td>
</tr>
</tbody>
</table>
11-use(s) activities to exchange information rather than grammar (Q13) | .184 | .206 | (n = 49) | 3.25 | 3.22 | -0.03
12-alter(s) speech for understanding (Q18) | .463 | .001* | (n = 47) | 3.36 | 3.34 | -0.02
13-demonstrate(s) native-like control in FL (Q15) | .112 | .453 | (n = 47) | 3.54 | 3.54 | 0.00
13-use(s) FL in the classroom (Q21) | -.063 | .666 | (n = 49) | 3.67 | 3.67 | 0.00

* significant at an alpha level of .05.

Only one item from the evaluation column of the Evaluation Questionnaire resulted in an averaged difference between individual teachers and their students of more than .30. Item 4 relating to the direct correction of oral mistakes resulted in a negative averaged difference across all teachers of -.529 meaning that, on average, individual teachers perceived their implementation of direct oral correction as less effective than did students. This was reflected in the overwhelming proportion of teachers (36/47) who felt their approach to direct oral correction was less effective than did their students, as compared to those who felt their approach was more effective than did students (9/47). The item with the next closest averaged teacher-student difference was Item 6 (has/have students complete specific tasks with real life purpose) with an averaged teacher-student difference of .288. The items with the smallest averaged teacher-student differences were Items 10 (address/es errors by explaining why), 13 (use/s activities to exchange information rather than grammar), 14 (present/s grammar in real-world context), 15 (demonstrate/s native-like control in FL), 18 (alter/s speech for understanding), and 21 (use/s FL in the classroom) all of which resulted in averaged differences of less than .10. Of these items, Item 15 concerning the importance of the teacher demonstrating native-
like control of the FL resulted in a minute averaged difference of .009 reflecting an apparent area of agreement between teachers and students.

Table 18. Summary of Averaged Differences between Individual Teachers’ Raw Scores and their Students’ Mean Responses and Direction of Difference by Item on Evaluation Column of Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>“In this class my foreign language teacher. . .”/ “As a foreign language teacher, I . . .”</th>
<th>Rank of Item</th>
<th>Item</th>
<th>Averaged Difference between Individual Teachers’ Raw Scores and their Students’ Mean Scores</th>
<th>Number of Teachers with Greater Perceived Effectiveness than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers with Lower Perceived Effectiveness than Students on Item and % of Responding Teachers</th>
<th>Number of Teachers Responding to Item and % of Total Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-correct(s) oral mistakes directly (Q4)</td>
<td>1</td>
<td>-.529</td>
<td>9 (20 %)</td>
<td>36 (80 %)</td>
<td>45 (%)</td>
<td></td>
</tr>
<tr>
<td>2-has/have students complete specific tasks with real life purpose (Q6)</td>
<td>2</td>
<td>.288</td>
<td>26 (60 %)</td>
<td>17 (40 %)</td>
<td>43 (%)</td>
<td></td>
</tr>
<tr>
<td>3-uses small-group work (Q12)</td>
<td>3</td>
<td>.255</td>
<td>30 (61 %)</td>
<td>19 (39 %)</td>
<td>49 (%)</td>
<td></td>
</tr>
<tr>
<td>4-grade(s) for grammatical accuracy (Q7)</td>
<td>4</td>
<td>-.250</td>
<td>15 (31 %)</td>
<td>34 (69 %)</td>
<td>49 (%)</td>
<td></td>
</tr>
<tr>
<td>5-use(s) computer-based technologies (Q1)</td>
<td>5</td>
<td>.245</td>
<td>18 (67 %)</td>
<td>9 (33 %)</td>
<td>27 (%)</td>
<td></td>
</tr>
<tr>
<td>6-has/have students use language outside of class (Q3)</td>
<td>6</td>
<td>-.162</td>
<td>14 (45 %)</td>
<td>17 (55 %)</td>
<td>31 (%)</td>
<td></td>
</tr>
<tr>
<td>7-encourage(s) students to speak the FL in class</td>
<td>7</td>
<td>-.149</td>
<td>17 (36 %)</td>
<td>30 (64 %)</td>
<td>47 (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Q11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-has/have students respond to physical commands—TPR (Q9)</td>
<td>.142</td>
<td>26 (57 %)</td>
<td>20 (43 %)</td>
<td>46 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-use(s) real-life materials in the classroom (Q17)</td>
<td>.142</td>
<td>27 (59 %)</td>
<td>19 (41 %)</td>
<td>46 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-assess(es) interaction with classmates (Q20)</td>
<td>-.134</td>
<td>18 (44 %)</td>
<td>23 (56 %)</td>
<td>41 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-teach(es) grammar with rules before examples (Q16)</td>
<td>-.132</td>
<td>15 (37 %)</td>
<td>26 (63 %)</td>
<td>41 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-devote(s) time for culture (Q2)</td>
<td>.131</td>
<td>24 (50 %)</td>
<td>24 (50 %)</td>
<td>48 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-engage(s) students in information gap activities (Q19)</td>
<td>.126</td>
<td>22 (50 %)</td>
<td>22 (50 %)</td>
<td>44 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-make(s) use of English (Q5)</td>
<td>-.119</td>
<td>14 (30 %)</td>
<td>33 (70 %)</td>
<td>47 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-correct(s) oral mistakes indirectly (Q8)</td>
<td>-.106</td>
<td>21 (43 %)</td>
<td>28 (57 %)</td>
<td>49 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-address(es) errors by explaining why (Q10)</td>
<td>-.082</td>
<td>23 (48 %)</td>
<td>25 (52 %)</td>
<td>48 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-present(s) grammar in real-world context (Q14)</td>
<td>.077</td>
<td>23 (47 %)</td>
<td>26 (53 %)</td>
<td>49 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-use(s) activities to exchange information rather than grammar (Q13)</td>
<td>.036</td>
<td>21 (43 %)</td>
<td>28 (57 %)</td>
<td>49 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-use(s) FL in the classroom</td>
<td>-.014</td>
<td>32 (65 %)</td>
<td>17 (35 %)</td>
<td>49 (%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When individual teacher-student differences are averaged across all 21 items on the evaluation column of the Evaluation Questionnaire, two teachers’ averaged differences with their students across all items was in excess of 1.00. Teacher 17 had a positive difference of 1.005 indicating, on average, a higher perceived effectiveness of implementation than students did, while Teacher 45’s averaged teacher-student difference resulted in a value of -1.150 reflecting a greater perceived effectiveness of the teacher’s behavior by students than by the teacher. Teacher 45 had an averaged difference with students of .906. On the other extreme lie Teachers 34, 38, and 49 all of whom had averaged differences less than .010. Table 19 displays each teacher’s averaged teacher-student difference across all 21 items on the evaluation column of the Evaluation Questionnaire including how the averaged difference ranks with the other teachers.

<table>
<thead>
<tr>
<th>Q21</th>
<th>.009</th>
<th>26 (55%)</th>
<th>21 (45%)</th>
<th>47 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19-demonstrate(s) native-like control in FL (Q15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-alter(s) speech for understanding (Q18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

208
Table 19. Averaged Teacher-Student Difference across 21 Items on Evaluation Column of Evaluation Questionnaire by Teacher.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across all 21 Items on Evaluation Column of Evaluation Questionnaire (4-point scale)</th>
<th>Rank</th>
<th>Teacher Number</th>
<th>Averaged Teacher-Student Difference Across all 21 Items on Evaluation Column of Evaluation Questionnaire (4-point scale)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.371</td>
<td>11</td>
<td>26</td>
<td>-.510</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>-.129</td>
<td>30</td>
<td>27</td>
<td>-.127</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>.270</td>
<td>19</td>
<td>28</td>
<td>-.118</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>.290</td>
<td>15</td>
<td>29</td>
<td>-.016</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>.256</td>
<td>20</td>
<td>30</td>
<td>-.133</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>-.446</td>
<td>8</td>
<td>31</td>
<td>-.597</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>-.406</td>
<td>10</td>
<td>32</td>
<td>.365</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>-.287</td>
<td>17</td>
<td>33</td>
<td>.186</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>-.190</td>
<td>24</td>
<td>34</td>
<td>.007</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>.369</td>
<td>13</td>
<td>35</td>
<td>-.127</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>.193</td>
<td>23</td>
<td>36</td>
<td>-.020</td>
<td>42</td>
</tr>
<tr>
<td>12</td>
<td>-.370</td>
<td>12</td>
<td>37</td>
<td>-.024</td>
<td>41</td>
</tr>
<tr>
<td>13</td>
<td>-.209</td>
<td>21</td>
<td>38</td>
<td>-.007</td>
<td>45</td>
</tr>
<tr>
<td>14</td>
<td>.039</td>
<td>40</td>
<td>39</td>
<td>-.171</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>-.017</td>
<td>43</td>
<td>40</td>
<td>.289</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>.539</td>
<td>6</td>
<td>41</td>
<td>-.066</td>
<td>36</td>
</tr>
<tr>
<td>17</td>
<td>1.005</td>
<td>2</td>
<td>42</td>
<td>-.283</td>
<td>18</td>
</tr>
<tr>
<td>18</td>
<td>0.116</td>
<td>33</td>
<td>43</td>
<td>.065</td>
<td>37</td>
</tr>
<tr>
<td>19</td>
<td>-.207</td>
<td>22</td>
<td>44</td>
<td>-.143</td>
<td>28</td>
</tr>
<tr>
<td>20</td>
<td>-.152</td>
<td>27</td>
<td>45</td>
<td>.906</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>-.100</td>
<td>35</td>
<td>46</td>
<td>.580</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>-.055</td>
<td>38</td>
<td>47</td>
<td>.109</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>.437</td>
<td>9</td>
<td>48</td>
<td>-1.150</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>.046</td>
<td>39</td>
<td>49</td>
<td>-.002</td>
<td>46</td>
</tr>
<tr>
<td>25</td>
<td>-.016</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Seven instructors with response patterns on the evaluation column of the Evaluation Questionnaire that correlated with their students’ to a degree significantly
different from zero were Teachers 2, 7, 12, 21, 25, 27, and 48. All seven teachers correlated positively with their students with Teacher 25’s responses resulting in the highest correlation with students’ at $r = .724$. A small minority of teachers had negative correlations, but they did not result significantly different from zero. Table 20 includes Pearson product-moment correlations for each individual teacher and their respective $p$ values as well as a column indicating how each teacher’s correlation ranks with the other teachers.

Table 20. Pearson Product-moment Correlation for Individual Teachers and their Students on Evaluation Column of Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Evaluation Column of Evaluation Questionnaire</th>
<th>Rank</th>
<th>$p$ value</th>
<th>Teacher Number</th>
<th>Pearson Product-moment Correlation for Individual Teachers on Evaluation Column of Evaluation Questionnaire</th>
<th>Rank</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.030</td>
<td>47</td>
<td>.905</td>
<td>26</td>
<td>.322</td>
<td>21</td>
<td>.192</td>
</tr>
<tr>
<td>2</td>
<td>.538</td>
<td>5</td>
<td>.012*</td>
<td>27</td>
<td>.488</td>
<td>9</td>
<td>.034*</td>
</tr>
<tr>
<td>3</td>
<td>.445</td>
<td>8</td>
<td>.056</td>
<td>28</td>
<td>.151</td>
<td>33</td>
<td>.513</td>
</tr>
<tr>
<td>4</td>
<td>.415</td>
<td>11</td>
<td>.077</td>
<td>29</td>
<td>.343</td>
<td>17</td>
<td>.163</td>
</tr>
<tr>
<td>5</td>
<td>-.142</td>
<td>35</td>
<td>.613</td>
<td>30</td>
<td>.340</td>
<td>18</td>
<td>.155</td>
</tr>
<tr>
<td>6</td>
<td>.265</td>
<td>22</td>
<td>.289</td>
<td>31</td>
<td>.423</td>
<td>10</td>
<td>.071</td>
</tr>
<tr>
<td>7</td>
<td>.364</td>
<td>15</td>
<td>.105</td>
<td>32</td>
<td>.084</td>
<td>39</td>
<td>.732</td>
</tr>
<tr>
<td>8</td>
<td>.604</td>
<td>3</td>
<td>.006*</td>
<td>33</td>
<td>.048</td>
<td>45</td>
<td>.856</td>
</tr>
<tr>
<td>9</td>
<td>.237</td>
<td>25</td>
<td>.301</td>
<td>34</td>
<td>.334</td>
<td>19</td>
<td>.139</td>
</tr>
<tr>
<td>10</td>
<td>.131</td>
<td>36</td>
<td>.581</td>
<td>35</td>
<td>.172</td>
<td>32</td>
<td>.482</td>
</tr>
<tr>
<td>11</td>
<td>.229</td>
<td>27</td>
<td>.331</td>
<td>36</td>
<td>.363</td>
<td>16</td>
<td>.116</td>
</tr>
<tr>
<td>12</td>
<td>.485</td>
<td>7</td>
<td>.041*</td>
<td>37</td>
<td>.217</td>
<td>30</td>
<td>.371</td>
</tr>
<tr>
<td>13</td>
<td>.067</td>
<td>42</td>
<td>.774</td>
<td>38</td>
<td>.051</td>
<td>44</td>
<td>.835</td>
</tr>
<tr>
<td>15</td>
<td>-.236</td>
<td>26</td>
<td>.330</td>
<td>40</td>
<td>-.075</td>
<td>40</td>
<td>.746</td>
</tr>
</tbody>
</table>
RESEARCH QUESTION 7

What impact do language type and class level exert on students’ responses to both the 24-item (Effective Teacher) and 21-item (Evaluation) questionnaires?

Two important independent variables that this research question addresses are language type and class level and their influence on student responses on each questionnaire. Since some of the languages represented in the study only had 1 or 2 classes participate, as was the case with Greek (1 class of 5 students), and Turkish (2 classes totaling 6 students), it was decided to group the languages. A frequently used distinction in the field of foreign language pedagogy refers to “commonly taught languages” (CTL) versus “less commonly taught languages” (LCTL). That distinction has been employed by Duff and Polio (1990) and Polio and Duff (1994) in their multi-language study into target language use. Therefore, all Spanish, French, and German classes were grouped together and classified as “commonly taught” (CTL) while Italian, Greek, Turkish, Hebrew, Japanese, and Arabic were deemed “less-commonly taught” (LCTL). Although the commonly taught languages, especially Spanish, outnumbered the less commonly taught languages, there were still enough participants for a meaningful
comparison to be made. Any further division of languages would have resulted in statistical analyses that would be very difficult to interpret with reduced statistical power.

The second independent variable that is included in this research question and the statistical models displayed in Tables 21, 22, and 23 has reference to level. All first-year courses, e.g., 101, 102, were grouped together, as were second-year courses, e.g., 201, 202. Some of the items on the questionnaires may demonstrate sensitivity to level as might be the case with items that refer to target language, such as the amount of English used in the classroom by the teacher, and the ideal time to begin requiring students to speak the language. The alpha level for all ANOVA models was adjusted from the commonly used .05 level to a Bonferroni adjusted level of .0021 for the 24-item Effective Teacher Questionnaire and .0024 for the 21-item Evaluation Questionnaire.

Tables 21, 22, and 23 include ANOVA models for only those items from each questionnaire where one of the three independent variables, i.e., language, level, and the interaction between language and level (interaction), produced a statistically significant effect on the dependent variable, i.e., students’ responses to each item. This means that a significant effect for language, for example, would mean that students’ responses on Item 1 on a given questionnaire are influenced by their status as a CTL or LCTL beyond mere chance. Table 21 includes the significant ANOVA results for the first survey, Effective Teacher Questionnaire, while Table 22 includes statistically significant effects by item for students’ responses to the frequency column of the Evaluation Questionnaire. Table 23 includes the significant ANOVA results from the evaluation column on the Evaluation Questionnaire.
When students and teachers were asked to reflect on their ideals of what a foreign language teacher should do (Effective Teacher Questionnaire), significant main effects for language and level resulted with no items demonstrating a significant effect for the interaction between language and level. Language resulted as the most frequent effect and appeared on 4 of the 24 items: 2 (assess group tasks), 4 (have students use language outside of class), 9 (be as knowledgeable about culture as language), and 10 (not grade production for accuracy). On Items 2 (assess group tasks) and 10 (not grade production for accuracy), students in CTL classes expressed greater agreement than their LCTL counterparts. The reverse was true for Items 4 (have students use language outside of class) and 9 (be as knowledgeable about culture as language) where LCTL students’ responses reflected greater agreement than CTL students’ responses.

The next most frequent statistically significant effect was detected for level with this variable resulting significant on only two items: 8 (only correct oral errors indirectly) and 14 (require students to speak FL first day of class). In both cases second-year students expressed greater agreement with Items 8 (only correct oral errors indirectly) and 14 (require students to speak FL first day of class) than did first-year students.
Table 21. ANOVA of Language Type and Class Level by Item for Students’ Responses to Effective Teacher Questionnaire by Item.

<table>
<thead>
<tr>
<th>Item Number “Effective foreign language teachers should . . .”</th>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 – assess group tasks</td>
<td>Language</td>
<td>1</td>
<td>9.59591384</td>
<td>24.25</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.03379387</td>
<td>.09</td>
<td>.7701</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.69570458</td>
<td>1.76</td>
<td>.1851</td>
</tr>
<tr>
<td>4 – have students use language outside of class</td>
<td>Language</td>
<td>1</td>
<td>6.53198010</td>
<td>10.26</td>
<td>.0014*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.04478027</td>
<td>.07</td>
<td>.7909</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.76372446</td>
<td>1.20</td>
<td>.2736</td>
</tr>
<tr>
<td>8 – only correct oral errors indirectly</td>
<td>Language</td>
<td>1</td>
<td>0.45143895</td>
<td>.82</td>
<td>.3664</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>5.24794929</td>
<td>9.49</td>
<td>.0021*</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.10775092</td>
<td>.19</td>
<td>.6590</td>
</tr>
<tr>
<td>9 – be as knowledgeable about culture as language</td>
<td>Language</td>
<td>1</td>
<td>11.74280732</td>
<td>22.72</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.77255521</td>
<td>1.49</td>
<td>.2217</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.02744861</td>
<td>.05</td>
<td>.8178</td>
</tr>
<tr>
<td>10 – not grade production for accuracy</td>
<td>Language</td>
<td>1</td>
<td>7.54575413</td>
<td>15.12</td>
<td>.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>1.77122316</td>
<td>3.55</td>
<td>.0598</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.51804150</td>
<td>1.04</td>
<td>.3085</td>
</tr>
<tr>
<td>14 – require students to speak FL first day of class</td>
<td>Language</td>
<td>1</td>
<td>5.59893771</td>
<td>8.53</td>
<td>.0036</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>7.53678090</td>
<td>11.48</td>
<td>.0007*</td>
</tr>
<tr>
<td>Interaction</td>
<td>1</td>
<td>0.03939274</td>
<td>.06</td>
<td>.8065</td>
<td></td>
</tr>
</tbody>
</table>

* significant at a Bonferroni adjusted alpha level of .0021 to avoid Type I error.

When students and teachers reflected on how often teachers performed the target teaching behaviors (frequency column of Evaluation Questionnaire), *language* again resulted as the most frequently occurring variable that resulted in a significant effect. Nine of the 24 items demonstrated significant effects for language and included the following items: 1 (uses computer-based technologies), 2 (devotes time for culture), 3 (has students use language outside of class), 4 (corrects oral mistakes directly), 7 (grades for grammatical accuracy), 8 (corrects oral mistakes indirectly), 9 (has students respond to physical commands—TPR), 12 (uses small-group work), and 14 (presents grammar in real-world context). Of these nine items with statistically significant effects for language type, only Items 1 (uses computer-based technologies), 2 (devotes time for culture), and 12 (uses small-group work) reflected greater perceived frequency among CTL students than LCTL students. The remaining six items all resulted in greater perceived frequency for LCTL students than CTL students: Items 3 (has students use language outside of class), 4 (corrects oral mistakes directly), 7 (grades for grammatical accuracy), 8 (corrects oral mistakes indirectly), 9 (has students respond to physical commands—TPR), and 14 (presents grammar in real-world context).

Unlike the first questionnaire, *level* and *interaction* each resulted in one item with a statistically significant effect on the dependent variable. Item 10 (addresses errors by explaining why) produced a significant effect for the variable level while Item 17 was statistically significant for the interaction between the two variables. First-year students
expressed greater perceived frequency than second-year students on Item 10 regarding teachers addressing errors by providing students with explanations. Item 17 (uses real-life materials in the classroom) presents an interesting interaction between the two variables. CTL students’ perceived level of frequency decreased from first to second year while the opposite was true for LCTL students as their perceived degree of frequency for teachers’ integration of real-life materials increased from first to second year.

Table 22. ANOVA of Language Type and Class Level by Item for Students’ Responses to Frequency Column on Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>“In this class my foreign language teacher...”</td>
<td>Language</td>
<td>1</td>
<td>15.97716704</td>
<td>12.28</td>
<td>.0005*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>3.85584136</td>
<td>2.96</td>
<td>.0854</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>4.47200455</td>
<td>3.44</td>
<td>.0640</td>
</tr>
<tr>
<td>1 – uses computer-based technologies</td>
<td>Language</td>
<td>1</td>
<td>47.39329462</td>
<td>132.90</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>1.28460873</td>
<td>3.90</td>
<td>.0580</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.08223996</td>
<td>.23</td>
<td>.6312</td>
</tr>
<tr>
<td>2 – devotes time for culture</td>
<td>Language</td>
<td>1</td>
<td>60.27351953</td>
<td>65.59</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.72564055</td>
<td>.79</td>
<td>.3744</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>3.32518025</td>
<td>3.62</td>
<td>.0574</td>
</tr>
<tr>
<td>3 – has students use language outside of class</td>
<td>Language</td>
<td>1</td>
<td>10.70502198</td>
<td>15.07</td>
<td>.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>3.07399014</td>
<td>4.33</td>
<td>0.0377</td>
</tr>
<tr>
<td>4 – corrects oral mistakes directly</td>
<td>Language</td>
<td>1</td>
<td>15.97716704</td>
<td>12.28</td>
<td>.0005*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>3.85584136</td>
<td>2.96</td>
<td>.0854</td>
</tr>
<tr>
<td>7 – grades for grammatical accuracy</td>
<td>Interaction</td>
<td>1</td>
<td>0.40414918</td>
<td>0.57</td>
<td>0.4508</td>
</tr>
<tr>
<td>8 – corrects oral mistakes indirectly</td>
<td>Language</td>
<td>1</td>
<td>7.08905420</td>
<td>20.50</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>9 – has students respond to physical commands—TPR</td>
<td>Level</td>
<td>1</td>
<td>0.77331218</td>
<td>2.24</td>
<td>.1351</td>
</tr>
<tr>
<td>10 – addresses errors by explaining why</td>
<td>Interaction</td>
<td>1</td>
<td>0.01821488</td>
<td>.05</td>
<td>.8185</td>
</tr>
<tr>
<td>12 – uses small-group work</td>
<td>Language</td>
<td>1</td>
<td>10.20386762</td>
<td>15.84</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Level</td>
<td>1</td>
<td>1.29868628</td>
<td>2.02</td>
<td>.1560</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Interaction</td>
<td>1</td>
<td>0.00835208</td>
<td>.01</td>
<td>.9094</td>
</tr>
<tr>
<td>9 – has students respond to physical commands—TPR</td>
<td>Language</td>
<td>1</td>
<td>14.57784871</td>
<td>12.98</td>
<td>.0003*</td>
</tr>
<tr>
<td>10 – addresses errors by explaining why</td>
<td>Level</td>
<td>1</td>
<td>0.34848211</td>
<td>.31</td>
<td>.5776</td>
</tr>
<tr>
<td>12 – uses small-group work</td>
<td>Interaction</td>
<td>1</td>
<td>7.15805536</td>
<td>6.37</td>
<td>.0117</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Language</td>
<td>1</td>
<td>4.49739427</td>
<td>6.91</td>
<td>.0087</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Level</td>
<td>1</td>
<td>7.25383553</td>
<td>11.15</td>
<td>.0009*</td>
</tr>
<tr>
<td>12 – uses small-group work</td>
<td>Interaction</td>
<td>1</td>
<td>0.25988954</td>
<td>.40</td>
<td>.5275</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Level</td>
<td>1</td>
<td>4.55993845</td>
<td>14.10</td>
<td>.0002*</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Interaction</td>
<td>1</td>
<td>0.00062568</td>
<td>.00</td>
<td>.9649</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Language</td>
<td>1</td>
<td>8.81691574</td>
<td>14.09</td>
<td>.0002*</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Level</td>
<td>1</td>
<td>0.02672785</td>
<td>.04</td>
<td>.8363</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Interaction</td>
<td>1</td>
<td>4.46708054</td>
<td>7.14</td>
<td>.0077</td>
</tr>
<tr>
<td>* significant at a Bonferroni adjusted alpha level of .0024 to avoid Type I error.</td>
<td>Language</td>
<td>1</td>
<td>1.04266930</td>
<td>1.25</td>
<td>.2644</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Level</td>
<td>1</td>
<td>0.87771718</td>
<td>1.05</td>
<td>.3058</td>
</tr>
<tr>
<td>17 – uses real-life materials in the classroom</td>
<td>Interaction</td>
<td>1</td>
<td>22.34023664</td>
<td>26.72</td>
<td>&lt;.0001*</td>
</tr>
</tbody>
</table>
Table 23 displays the ANOVA results for the evaluation column on the Evaluation Questionnaire which asked students to reflect on how effective teachers concretely implemented each teaching practice in their respective classrooms if, indeed, they did. The independent variable language again proved to be the most common significant effect demonstrating a statistically significant influence on five items: 3 (has students use language outside of class), 11 (encourages students to speak the FL in class), 14 (presents grammar in real-world context), 19 (engages students in information gap activities), and 21 (use/s FL in the classroom). In every case LCTL students expressed greater perceived effectiveness by their teachers with each of the five items than did CTL students. The only other variable of the three that produced a statistically significant effect was level which occurred on Item 1 (uses computer-based technologies). In this case first-year students perceived greater effectiveness of their teachers’ use of computer-based technologies than second-year students.
Table 23. ANOVA of Language Type and Class Level by Item for Students’ Responses to Evaluation Column on Evaluation Questionnaire.

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – uses computer-based technologies</td>
<td>Language</td>
<td>1</td>
<td>0.85352974</td>
<td>1.28</td>
<td>.2579</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>11.21710762</td>
<td>16.85</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>2.28229005</td>
<td>3.43</td>
<td>.0645</td>
</tr>
<tr>
<td>3 – has students use language outside of class</td>
<td>Language</td>
<td>1</td>
<td>24.54248487</td>
<td>31.12</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.81554200</td>
<td>1.03</td>
<td>.3095</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.15004439</td>
<td>.19</td>
<td>.6628</td>
</tr>
<tr>
<td>11 – encourages students to speak the FL in class</td>
<td>Language</td>
<td>1</td>
<td>6.23780543</td>
<td>10.69</td>
<td>.0011*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.07950369</td>
<td>.14</td>
<td>.7121</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.23305868</td>
<td>.40</td>
<td>.5275</td>
</tr>
<tr>
<td>14 – presents grammar in real-world context</td>
<td>Language</td>
<td>1</td>
<td>6.52599885</td>
<td>12.67</td>
<td>.0004*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.00225901</td>
<td>.00</td>
<td>.9472</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>1.07130137</td>
<td>2.08</td>
<td>.1496</td>
</tr>
<tr>
<td>19 – engages students in information gap activities</td>
<td>Language</td>
<td>1</td>
<td>9.44408401</td>
<td>15.08</td>
<td>.0001*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>1.11251811</td>
<td>1.78</td>
<td>.1829</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1</td>
<td>0.87428189</td>
<td>1.40</td>
<td>.2377</td>
</tr>
<tr>
<td>21 – uses FL in the classroom</td>
<td>Language</td>
<td>1</td>
<td>4.17348525</td>
<td>13.18</td>
<td>.0003*</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>1</td>
<td>0.05193494</td>
<td>.16</td>
<td>.6856</td>
</tr>
</tbody>
</table>
RESEARCH QUESTION 8

*How do students’ evaluations of their foreign language teachers on the 21-item language-specific instrument (Evaluation Questionnaire) coincide with or differ from their evaluation of their teachers on selected items from the general, university-wide evaluation form, i.e., the Teacher-Course Evaluation (TCE)?*

Research Question 8 compares students’ evaluations of teachers across all 21 items on the *Evaluation Questionnaire* with students’ responses to two questions taken directly from the TCE. Six total questions were taken from the TCE and reproduced on this study’s instrumentation, including the following: 1) What is your overall rating of this instructor’s teaching effectiveness?, 2) How much do you feel you have learned in this course?, 3) What is your overall rating of this course?, 4) Rate the usefulness of the in-class activities (lectures, discussions, etc.) in this course in helping you learn?, 5) I was treated with respect in this class, 6) What is your rating of this instructor compared with other instructors you have had? After careful analysis of all six of these items, it was determined that only two of the questions, the first and the sixth, related directly to the teacher and teachers’ behaviors leading to those two being correlated to the discipline-specific instrument used in the current study.

As mentioned previously, this comparison should be considered exploratory and tentative as the items on the Evaluation Questionnaire cover a specific, limited set of teaching practices in the foreign language classroom on a 4-point scale whereas the TCE
questions are not specific to FL classrooms and most likely include other constructs relative to effective teaching, regardless of discipline, on a 5-point scale. In order to provide a statistically valid comparison between the two selected TCE questions and the Evaluation Questionnaire given to participants, two separate Pearson product-moment correlations were run to ascertain relationships between the two instruments. The first correlation examines the relationship between the dependent variable, TCE1, which is comprised of students’ means on the first TCE question which simply states “What is your overall rating of this instructor’s teaching effectiveness?” and the independent variable, students’ means on the evaluation column of the Evaluation Questionnaire (Evaluation). The second correlation includes the same student data from the Evaluation Questionnaire along with students’ responses to TCE Question 6, TCE6, which is the only other TCE question included that relates directly to the teacher and teaching behaviors—“What is your rating of this instructor compared with other instructors you have had?”

In both correlations, the relationship between students’ means on the Evaluation Questionnaire and students’ means on the TCE resulted in a significant correlation. Students’ means on the evaluation instrument used in this study significantly predicted a portion of the variance in the TCE scores. In the first correlation conducted, students’ mean responses on the Evaluation Questionnaire accurately predicted half of the variance in TCE Question 1 ($r^2 = .4944$). Students’ responses to the first TCE question regarding overall effectiveness correlated at a rate of .703 to the mean of students’ responses to the evaluation column on the Evaluation Questionnaire. That is to say that the dependent
variable, students’ responses to the TCE question stating “What is your overall rating of this instructor’s teaching effectiveness,” demonstrated a statistically significant correlation ($r = .703$) with the independent variable, students’ mean responses on all 21 items of the Evaluation Questionnaire. The resulting $p$ value for the correlation between TCE1 and the students’ means from the evaluation column of the Evaluation Questionnaire was .0001 indicating that the correlation was significantly different from zero.

The second Pearson product-moment correlation also produced a statistically significant correlation ($p< .0001$) between the independent variable, students’ means from the 21 language-specific evaluation items, and the dependent variable, students’ means from TCE Question 6 which states “What is your rating of this instructor compared with other instructors you have had?” This correlation analysis explained less than half of the variance ($r^2 = .4333$) of the dependent variable with a correlation between the independent and dependent variables of $r = .658$. In summary, the mean of students’ responses to the instrument designed and administered by this researcher demonstrated a significant correlation to TCE Question 1 and TCE Question 6 which relates specifically to teachers and their effectiveness. Hence, one could predict to a significant degree beyond chance students’ responses on TCE Question 1 and 6 by how they responded on the evaluative instrument prepared for this study.
RESEARCH QUESTION 9

What relationship exists between students’ response patterns on both the Effective Teacher and Evaluation Questionnaires and students’ responses to selected items from the TCE instrument?

The final research question of this large-scale study delves into the relationship between students’ evaluations of their teachers on the TCE questions, specifically Question 1 (“What is your overall rating of this instructor’s teaching effectiveness?”) and Question 6 (“What is your rating of this instructor compared with other instructors you have had?”), and their responses on the Effective Teacher and Evaluation Questionnaires included in the current study. The intent of this research question was to determine whether students’ responses on the current study’s instrumentation demonstrated a significant correlation with students’ evaluations on the two specific TCE questions.

Following a similar protocol adopted in Research Question 8, the first question from the TCE was considered the dependent variable in the first multiple regression analysis and in the second model the sixth TCE question included in the study was the dependent variable. The independent variables, then, would be the means of students’ responses on the Effective Teacher Questionnaire (Effective), the frequency column (Freq.) and the evaluation column (Evaluation) of the Evaluation Questionnaire. In both cases the amount of variance explained by the multiple regression model was 50% or less, i.e., $r^2 = .4997$ for TCE Question 1 and $r^2 = .4355$ for TCE Question 6. Each model’s respective multiple correlation was relatively high even with three independent variables, $r = .707$ for TCE Question 1 and $r = .660$ for Question 6. While the first model (TCE
Question 1) resulted in statistically significant relationships between all independent variables, i.e., students’ responses on each questionnaire, and the dependent variable, the second model (TCE Question 6) had two of the three independent variables demonstrate a significant relationship with the dependent variable. The frequency column of the Evaluation Questionnaire did not maintain a statistically significant relationship with TCE Question 6. In summary, students’ responses on each questionnaire significantly predicted how students evaluated their teachers on TCE Question 1. In the case of students’ responses on TCE Question 6, only the frequency column of the Evaluation Questionnaire did not show a significant relationship in the regression model. Tables 24, 25, 26, and 27 present the results of both multiple regression models with TCE1 as the dependent variable in the first and TCE6 as the dependent variable in the second.

Table 24. Multiple Regression Analysis of TCE Question 1 and Students’ Responses on All Questionnaires.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>143.06313</td>
<td>390.16</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1172</td>
<td>.36667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1175</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significant at an alpha level of .05.

Table 25. Parameter Estimates for Multiple Regression Analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Degrees of</th>
<th>Parameter</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Degrees of Freedom</td>
<td>Mean Square</td>
<td>F value</td>
<td>p value</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>197.44935</td>
<td>301.36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1172</td>
<td>.65520</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1175</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant at an alpha level of .05.

Table 26. Multiple Regression Analysis of Mean of TCE Question 6 and Students’ Responses on All Questionnaires.

| Parameter Estimates
<table>
<thead>
<tr>
<th>Variable</th>
<th>Degrees of Freedom</th>
<th>Parameter Estimate</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-.42702</td>
<td>-2.88</td>
<td>.0040</td>
</tr>
<tr>
<td>Effective</td>
<td>1</td>
<td>-.21591</td>
<td>-2.01</td>
<td>.0447</td>
</tr>
<tr>
<td>Freq.</td>
<td>1</td>
<td>-.63483</td>
<td>-.55</td>
<td>.5791</td>
</tr>
<tr>
<td>Evaluation</td>
<td>1</td>
<td>3.91584</td>
<td>3.45</td>
<td>.0006</td>
</tr>
</tbody>
</table>

* Significant at an alpha level of .05.
SUMMARY

This chapter has presented the results of this study by answering each research question in turn and by summarizing the results in tables and figures. Overall, the most striking difference that surfaced from these results relates to the overwhelming degree of disparity between students’ and teachers’ perceptions on both questionnaires, but especially the Effective Teacher Questionnaire, on issues such as grammar teaching, error correction, and target and native language use. When averaged across all teachers and all students, teachers’ responses were more varied, reflective of stronger opinions regarding the target teaching behaviors, while students’ opinions were more mild and less polarized. Of particular interest is the apparent desire on the part of students to engage in more grammar practice as compared to communicative practice. Also, students seemed more interested in direct error correction of oral mistakes than did their teachers. Teachers appeared more convinced of the value of teaching culture and of communicative approaches to language teaching such as task-based teaching than their students.

Chapter 5 will synthesize the findings presented in Chapter 4 in more depth, commenting on trends in teacher-student differences for each category. Furthermore, Chapter 5 will present the pedagogical implications from the findings, the limitations of the current study and its instrumentation, and recommendations for future research.
CHAPTER 5
DISCUSSION AND CONCLUSIONS

INTRODUCTION

This fifth and final chapter will interpret the results presented in Chapter 4 and discuss theoretical and pedagogical implications as well as limitations of the current study. In addition, Chapter 5 will offer directions for future research in the area of students’ and teachers’ perceptions of foreign language teaching and their evaluations.

One of the most challenging aspects of a large-scale, questionnaire study is to meaningfully summarize and interpret the results so that certain findings are not over-embellished while others are neglected. With 49 participating teachers and their 1400 students from 9 different languages who completed two 20+-item questionnaires, the number of possible comparisons is seemingly infinite.

In spite of its character as an exploratory study, an attempt has been made by the researcher to analyze and present the results in such a way as to highlight those teacher-student comparisons that seem to uncover trends reflective of the sample population. Therefore, the following section will serve as a summary of the results as considered by category and item rather than by research question. At times it may appear to parallel Chapter 4’s approach by citing specific findings; nevertheless, Chapter 5 approaches the results by topic, allowing for a more coherent interpretation of the findings as they relate to specific practices. Appropriate implications and conclusions will accompany this thematic review of the results. Additionally, assessment implications taken from the comparisons between selected TCE questions and this study’s instrumentation will be
addressed, along with a discussion focusing on the relationship between teacher demographics and teacher-student matches and mismatches. Finally, this chapter will conclude by addressing pedagogical implications as well as limitations of the study, and directions for future research.

IMPLICATIONS OF RESULTS BY CATEGORY AND QUESTIONNAIRE

The discussion of each category presented in this section will be divided into two parts: 1) teachers’ and students’ responses on the Effective Teacher Questionnaire where participants reflected on “effective foreign language teachers” in general, and, 2) teachers and students responses on the Evaluation Questionnaire where each group reflected on their individual classes and teachers.

**Category:** Grammar Teaching; **Questionnaire:** Effective Teacher

When students and teachers reflected on whether they felt an effective FL teacher should *not* assess language production for accuracy (Item 10), teachers were slightly more in agreement than students, but not to a significant degree when a t-test was run comparing all students’ scores to all teachers’ scores. Mean student scores on the 4-point Likert scale were 2.66 and 2.98 for teachers. The average difference between individual teachers and their students was moderate at .293 but an overwhelming majority of teachers were in more agreement than their students on the item—34 of 49 total participating teachers.

One of the most striking discrepancies between teachers’ and students’ perceptions regarding grammar teaching surfaced on Item 16 on the Effective Teacher Questionnaire where a significant difference resulted from the t-test comparing all
teachers and all students as well as the second largest averaged teacher-student difference across all 24 items, -.694. Thirty-seven of 49 teachers were in greater disagreement than their students that “Effective foreign language teachers should mostly use activities that practice specific grammar points rather than activities whose goal is to merely exchange information.” This means that not only did a high averaged difference exist between individual teachers and their students but nearly all of the disagreement was in the same direction—students agreeing with the prompt statement more than their teachers. Clearly, teachers seem to value, or at least report that they value, the exchange of information more than practicing specific grammar points, and conversely, students do not consider the exchange of information nearly as valuable as their teachers.

Students and teachers expressed similar opinions regarding the use of grammar in a real-world context (Item 18) and the use of inductive approaches to grammar teaching (Item 20). Both groups agreed to almost the same degree that effective foreign language teachers should use grammar in a real-world context (Item 18). No significant difference resulted in the overall t-test comparing all teachers and all students and averaged teacher-student differences were very low. Participants’ opinions concerning inductive approaches to grammar teaching (Item 20) were quite neutral overall and did not result in statistically significant differences as to whether an effective teacher should present examples before rules. Similarly, averaged teacher-student differences across all 49 teachers were rather small at .057 on a 4-point scale.
**Category:** Grammar Teaching; **Questionnaire:** Evaluation

As no t-tests were run on the Evaluation Questionnaire due to the differing nature of the prompt statement for teachers and students, all analyses of responses to the frequency and evaluation column of the Evaluation Questionnaire will be reported using descriptive statistics—even in comparing teachers’ and students’ responses overall. A correlation analysis was run to determine how all teachers’ responses and all students’ responses correlated on each item and discussion of those results will be included. In reflecting on their specific classes and how often they graded for grammatical accuracy (Item 7), a correlation significantly different from zero resulted between all students’ and all teachers’ responses with overall teacher-student differences being minimal. Both teachers and students’ responses approached the maximum value ‘4’, or “frequently”. In evaluating their teachers’ effectiveness in grading for grammatical accuracy minimal raw differences resulted, overall and averaged by individual teacher. However, a large majority of teachers perceived their effectiveness in grading grammatical accuracy as lower than did their students (34/49).

Overall, teachers and students perceived teachers to be teaching grammar in real-world contexts (Item 14) and perceived their concrete execution of this technique as effective. A significant correlation between all students’ and all teachers’ responses resulted when reflecting on the frequency of occurrence but not on the effectiveness of the teacher’s implementation. Very minimal score differences resulted overall and at the averaged individual teacher level for both columns.
When asked how often participating teachers used deductive grammar teaching (Item 16), teachers and students coincided in their responses to the frequency column as they resulted in significant correlations. In spite of this correlation, the averaged difference between individual teachers and their students was moderate with a value of -.377 indicating that teachers perceived deductive grammar less frequently—28 of 47 responding teachers perceived less frequency than their students. The evaluation column resulted in both teachers and students perceiving teachers’ integration of deductive grammar as effective with minimal differences overall and averaged across individual teachers and their students.

**Category**: Error Correction; **Questionnaire**: Effective Teacher

The item that provoked the strongest and most divisive response on either questionnaire relates to error correction and states that effective foreign language teachers should *not* correct students immediately after making a mistake in speaking (Item 5). Across all teachers and all students the comparison resulted in almost an entire point difference on a 4-point scale with students’ means falling on the disagree side of neutral and teachers falling on the agree side. These values resulted in statistically significant differences when all students’ and all teachers’ responses were compared using a t-test. This item on error correction resulted in the largest averaged difference across all 24 items between individual students and their teachers with a value of .897. Almost 80% of teachers (39/49) expressed greater agreement than did their students on this item. In general, the teachers from this study feel much more hesitant to correct students’ oral mistakes immediately than do students.
As a group, teachers agreed a bit more strongly than students that effective foreign language teachers should use indirect means of correcting students’ oral errors rather than direct means (Item 8). Averaged differences between individual teachers and students resulted in a moderate value of .233 on a 4-point scale. Nevertheless, a majority of teachers (30/49) agreed with this item more strongly than did their respective students.

Students’ responses overall to Item 13 resulted in statistically significant differences from teachers’ with students expressing greater agreement that teachers should provide explanations following error correction. The average of individual teacher’s differences with students resulted in a difference of -.427 with 41 of 49 teachers disagreeing more strongly than their students that effective foreign language teachers should address errors with immediate explanations.

**Category:** Error Correction; **Questionnaire:** Evaluation

When students and teachers were asked how often they observed *direct* oral correction (Item 4), a significant correlation was found overall on the item although averaged individual teacher-student differences were -.478 with 39 of 49 teachers perceiving lower frequency than their students. On the contrary, when teachers and students were asked how often teachers corrected oral mistakes *indirectly* (Item 8), teachers perceived themselves to be using this correction strategy more often than did students. Thirty-six of 49 teachers perceived themselves to be correcting oral mistakes more often than their students with an averaged teacher-student difference across all teachers and their respective students of .357. Students’ and teachers’ evaluation of teachers’ use of this error correction technique did not result in a significant correlation or
large differences. Both seemed to feel that teachers were somewhat effective in correcting oral mistakes indirectly. Participants’ responses to Item 10 (address/es errors by explaining why) did not correlate significantly and reflected very minimal differences on both the frequency column and the evaluation column of the Evaluation Questionnaire. Teachers as well as students seemed to perceive teachers as addressing errors by explaining why quite frequently and quite effectively.

**Category:** Target Language Use; **Questionnaire:** Effective Teacher

When considered as two separate groups, both teachers and students seemed undecided when asked whether effective foreign language teachers should *not* use English in the classroom (Item 7) with mean scores around 2.5, or neutral, on a scale from 1-4. When their mean responses are compared on this item no statistically significant difference resulted. An averaged teacher-student difference across all teachers resulted in a value difference of .328. No clear pattern resulted from the number of teachers with greater agreement than students as 29 agreed with the item more than their students while 20 disagreed with it more than their students.

Three other items relative to target language use resulted in minimal differences between teachers and students, including Items 17 (ask students to speak FL only when ready), 19 (speak with native-like control of language), and 22 (not simplify speech). Items 17 and 22 appeared to spur neutral responses with neither group agreeing, or disagreeing, strongly with the items. While Item 19 led to minimal differences between the two groups similar to Items 17 and 22, this item resulted in both expressing strong
agreement, with a mean response across all students of 3.25 and a mean across all teachers of 3.12 on a 1-4 point scale.

One area of teacher-student difference that surfaced from the data related to whether teachers should require students to speak the FL the first day of class (Item 14). A statistically significant difference resulted when students’ and teachers’ responses were compared overall, with teachers in greater agreement that indeed students should speak the FL the first day of class. Similarly, this item produced the third largest averaged difference between individual teachers and their students with an averaged difference value of .623. Thirty-eight of the 49 responding teachers were individually more in agreement than their students that students should speak the FL the first day of class.

**Category:** Target Language Use; **Questionnaire:** Evaluation

Item 5 required students to reflect on the use of English in the classroom by the teacher, specifically the frequency of use and the teacher’s effectiveness in using it. A significant correlation resulted overall between teachers and students when reflecting on the frequency of use, but not on the effectiveness of English use. Minimal teacher-student differences resulted overall and when averaged by teacher on both the frequency and evaluation columns on this item. Item 21 (use/s FL in the classroom) closely relates to Item 5’s question about language, but in this instance students and teachers were asked to reflect on teachers’ use of the foreign language in the classroom. Unlike Item 5, no significant correlations resulted between students’ and teachers’ responses on either column of the Evaluation Questionnaire. Averaged individual teacher-student differences resulted minimal (<.053) on both the frequency and evaluation columns of the Evaluation
Questionnaire. An additional question posed to both groups relative to FL use refers to whether teachers demonstrate native-like control of the target language. Teachers’ and students’ responses to neither the frequency column nor the evaluation column correlated significantly. Averaged teacher-student differences on the item were minimal across both columns, less than .15, although 40 of 48 teachers had students who perceived their teacher’s control of the language to be native-like more often than did teachers themselves. Clearly, this item was designed with the teacher’s perspectives in mind as students’ ability to detect native-like control of the language might be difficult.

When students and teachers were asked to report on how frequently and how effective teachers encouraged students to speak the foreign language in class (Item 11), a significant correlation resulted overall across all teachers and all students on the frequency column, but not on the evaluation column. It is interesting to note that 36 of 47 teachers perceived themselves to be encouraging students to speak the FL in class more frequently than students; however, when both groups evaluated how effective the teachers were in encouraging their students, 30 of 47 teachers perceived lower effectiveness than their students in encouraging students.

Concerning the altering of language by teachers to facilitate student understanding, a significant correlation resulted between teachers and students on the evaluation column, not the frequency column, of the Evaluation Questionnaire. While averaged teacher-student differences across all teachers were moderate, .239, for the frequency column they were extremely small for the evaluation column, .009. On both
the frequency and evaluation columns more teachers perceived greater frequency and greater effectiveness than students.

**Category:** Culture; **Questionnaire:** Effective Teacher

Teachers appeared more interested in and desirous of integrating culture in the foreign language classroom than did students. When asked whether effective foreign language teachers should devote as much time to the teaching of culture as to the teaching of language (Item 3), teachers expressed stronger agreement than students that resulted in a statistically significant difference. This difference was reflected in the averaged teacher-student differences taken from each teacher with a result of .528. Once again, a large majority of teachers (39/49) expressed stronger agreement on this item than their students.

In responding to a question regarding the degree of cultural knowledge that an effective teacher must possess (Item 9), teachers on the whole were in more agreement than their students that they should be at least as knowledgeable about the culture as they are about the language. This difference resulted in a statistically significant difference overall and an averaged difference of .420 with most teachers agreeing more strongly with the item than their students. The item relative to the use of real-life materials by effective teachers (Item 21) resulted in both groups overall expressing agreement with no statistically significant difference. Likewise, averaged individual teacher-student discrepancies were minimal.
**Category:** Culture; **Questionnaire:** Evaluation

The issue of how much time to devote to culture in the FL classroom (Item 2) resulted in a significant correlation between teachers’ and students’ responses on the frequency column but not on the evaluation column. Averaged teacher-student differences across all teachers were less than .20. Nevertheless, 38 of 48 teachers perceived that they devoted time for culture more frequently than did their students.

In regard to the use of real-life materials in the FL classroom (Item 17) the comparison between all students’ and all teachers’ responses produced a significant correlation on the frequency column but not on the evaluation column. When individual teacher-student differences were averaged across all individual teachers on both columns, moderate differences resulted (frequency column—.222; evaluation column—.142). In both cases more than half of all teachers perceived themselves to use real-life materials with higher frequency and effectiveness than their students.

**Category:** Computer-based Technology; **Questionnaire:** Effective Teacher

When students’ and teachers’ responses were compared overall, teachers agreed more strongly that the frequent use of computer-based technologies was reflective of an effective foreign language teacher (Item 1). This difference resulted in a statistically significant difference between teachers and students. The averaged difference between individual teachers and their students was .422 with 36/49 of them expressing greater agreement than their students on the item.
**Category:** Computer-based Technology; **Questionnaire:** Evaluation

In reality, only 27 teachers self-evaluated in this area, implying that the remaining 22 teachers “Never” used computer-based technologies in the classroom. While students perceived a slightly higher rate of computer use than their teachers overall, most teachers who did use computers felt they used them more effectively than their students did, 18 of 27 to be exact.

In many cases the use of computers in the classroom and outside of the classroom may be restricted by such things as availability and imposed curricula. The sample of teachers who participated in this study, apparently, are privy to the recent push to integrate computer-based technologies in FL teaching. Nevertheless, these teachers, possibly for reasons mentioned above, have not concretely implemented the computer in their teaching of the foreign language in the classroom. Frequent incorporation of the computer in classroom foreign language teaching, has not yet, it seems, found its way into the courses that participated in this study.

**Category:** Communicative Language Teaching Strategies; **Questionnaire:**

Effective Teacher

Overall, and by teacher, teachers were more in agreement that effective foreign language teachers should require students to use the foreign language outside of class (Item 4). The overall comparison between teachers and students resulted in statistically significant differences and the averaged individual teacher-student values differed from students at a raw value of .465 with 35 of 49 teachers agreeing more than their students with the item. One of the goals of the National Standards in Foreign Language Education
Project (1999) to take the language outside of the classroom appears to be supported more fervently by teachers than students.

In regard to task-based teaching (Item 11) teachers’ and students’ responses overall resulted in significant differences after a t-test was conducted. When individual teacher-student differences were averaged a mean difference of .546 resulted—the fourth largest across all 24 items. An overwhelming majority of teachers, 36 of 48, agreed more than their students that effective foreign language teachers should use tasks rather than grammar-focused exercises in teaching the language. Asher’s TPR approach to foreign language teaching (Item 12) elicited more neutral responses with overall teacher and student means approaching significance but falling just shy of the Bonferroni adjusted alpha level of .0021. Averaged teacher-student differences produced a value of .297 with just over half of responding teachers (27/48) expressing greater agreement than their students that effective foreign language students should have students respond to commands physically in the foreign language.

As noted previously, teachers seemed much more convinced of the need for foreign language students to participate in a meaningful exchange of information during class. When asked about information-gap activities (Item 24), or activities where students have to find out unknown information from classmates using the foreign language, teachers’ responses resulted in statistically significant differences from students in the direction of stronger agreement by teachers. At the individual level, the averaged difference between teachers and their students was .476 with a large majority of teachers expressing greater agreement with the item than their students (34/49).
The often used strategy of putting foreign language students in groups to maximize exposure to and production of the foreign language (Item 15) resulted in statistically significant differences between teachers and students overall. Forty-one of 49 teachers disagreed more strongly than their students to the negatively worded item which reads “Effective foreign language teachers should *not* use predominantly small groups or pair work to complete activities in class.” The averaged teacher-student difference across teachers was -.510 where a negative indicates greater disagreement by teachers than students.

**Category**: Communicative Language Teaching Strategies; **Questionnaire**: Evaluation

Task-based teaching (Item 6) did not result in significant correlations between teachers and students on the frequency or evaluation columns. An averaged teacher-student difference across all teachers resulted in a positive difference of .337 with 31 of 49 teachers perceiving greater frequency than their students on this item. The evaluation column produced minimal differences.

Teachers appeared to feel that they used Total Physical Response (TPR) more frequently and more effectively than students. A significant correlation between teachers and students resulted on the evaluation column of the Evaluation Questionnaire although a small averaged teacher-student difference of .142 resulted. In regard to frequency, teachers differed from their students, on average, at .422 with 32 of 49 individual teachers perceiving greater frequency than did students.
Significant correlations were produced when teachers’ and students’ perceptions of small group work were compared on the frequency column with both groups perceiving very frequent use with means of 3.76 (teachers) and 3.73 (students) on a 4-point scale. Although averaged teacher-student differences across teachers was minimal (.033) an overwhelming majority (37/49) of teachers perceived greater frequency than did their students. This averaged difference increased (.255) when individual teachers and their students reflected on how effective teachers integrated group work. Nevertheless, a large proportion of teachers 30/49 perceived greater effectiveness of their use of group work than did their students.

Item 13, relating to the use of activities to exchange information rather than practice grammar, correlated significantly between teachers and students on the frequency column but not on the evaluation column. Averaged teacher-student differences on both columns were less than .10. A similar item, Item 19, questions how often and how effective teachers use information gap activities, or activities where students must exchange unknown information from partners. A significant correlation resulted only on the frequency column and minimal averaged teacher-student differences (<.150) for individual teachers resulted on both columns of the Evaluation Questionnaire.

**Category:** Assessment; **Questionnaire:** Effective Teacher

When students and teachers responded to Item 2, “Effective foreign language teachers should base at least some part of students’ grades on completion of assigned group tasks”, teachers agreed more strongly than students, with the difference resulting statistically significant. However, it is interesting to note that this was the only item of all
24 items on the Effective Teacher Questionnaire where a significant correlation existed between all teachers’ and all students’ responses. Most individual teachers were in greater agreement than their students with 37 of 49 reporting greater agreement than students that teachers should base at least some part of students’ grades on the completion of assigned group tasks. The averaged individual teacher-student difference was .492 on the 4-point scale.

On Item 6 on the Effective Teacher Questionnaire, students and teachers were asked whether effective foreign language teachers should allow students to respond to test questions in listening and reading via English rather than the foreign language. Overall, students’ and teachers’ responses were quite similar, as both groups’ means approached 2.5, or neutral, resulting in no statistically significant difference. That is to say that neither group was truly in agreement or disagreement with the use of English to test listening and reading; in fact, both groups’ means were mathematically in disagreement as they both fell short of 2.5 on a 1-4.0 scale, with 1 representing strong disagreement. The averaged individual teacher-student difference was minimal with a difference of .081.

The evaluation of grammatical accuracy was taken up on Item 10 which states that effective foreign language teachers should not grade language production (i.e., speaking and writing) primarily for grammatical accuracy. Teachers were, overall, slightly more in agreement with this statement than students, although that difference did not result to be statistically significant. On average, individual teachers differed from their students with a value of .293 with 34 of 48 responding teachers in greater agreement
than their students. Similarly, significant differences did not result between teachers and students responses overall when both groups were asked whether effective foreign language teachers should assess students’ ability to successfully interact with classmates (Item 23). Averaged individual teacher-student differences were moderate with a value of .237. A majority of teachers (33/49) expressed stronger agreement than their students on this item.

**Category**: Assessment; **Questionnaire**: Evaluation

Of the two items relative to assessment (Items 7 and 20), the frequency column of Item 7 (grade/s for grammatical accuracy) was the only one where a significant correlation resulted. Averaged teacher-student differences were minimal on the frequency column for Item 7 (-.120) but increased to -.250 on the evaluation column with 34 of 49 responding teachers perceiving lower effectiveness than their students on how they graded for grammatical accuracy.

Item 20 probes into the assessment of interaction with classmates, presenting almost the reverse situation to Item 7’s focus on grading for grammatical accuracy. Both groups overall means were somewhat neutral, 2.60 for teachers and 2.81 for students, and averaged teacher-student differences decreased from the frequency column to the effectiveness column from -.204 to -.134. In contemplating on the frequency of grading interactions with classmates and teachers’ effectiveness in doing so, a majority of teachers perceived lower frequency and effectiveness than did their students.
TCE COMPARISONS

The comparisons drawn between the official questions on the Teacher-Course Evaluations administered at the University of Arizona and the Evaluation Questionnaire administered as part of this study resulted in somewhat intuitive findings: students ratings of their teachers on the selected TCE questions paralleled their ratings on the subject-specific instrumentation designed for this study. Students responded to six total TCE questions in the current study, but only 2 of them were analyzed statistically as they were deemed to be more directly related to the teacher and teaching behaviors within the control of the teacher than the other four. In the first case, the mean from TCE Question 1, relative to students’ perceptions of their teacher’s overall effectiveness was compared to the mean of students’ responses to the evaluation column of the Evaluation Questionnaire. Second, the mean of TCE Question 6 which asked students to compare their teacher to other teachers was compared to the same set of student responses from the Evaluation Questionnaire. In both cases, the multiple regression analysis produced significant relationships between the two TCE questions and the Evaluation Questionnaire. Among the two comparisons, students’ responses from the Evaluation Questionnaire explained more of the statistical variance in TCE Question 1 than in TCE Question 6 and, likewise, led to a higher correlation with TCE Question 1 than TCE Question 6.

It is interesting to note that although both instruments, the TCE and the Evaluation instrument, purport to measure a number of varied teaching constructs they still demonstrate significant correlations (.703 for TCE Question 1 and .658 for TCE
Question 6). The Evaluation Questionnaire used in this study includes many issues specifically related to language teaching (i.e., target language use, grammar teaching, communicative language teaching) while the two TCE questions are broad, not bound by discipline, such as “What is your overall rating of this instructor’s teaching effectiveness?” and “What is your rating of this instructor compared with other instructors you have had?” However, it may be that these broad questions elicit discipline-specific responses from students.

Somewhat intuitive is the greater explanatory power of students’ responses from the Evaluation Questionnaire on TCE Question 1 relative to overall effectiveness as compared to TCE Question 6 which asks for a comparison between students’ current teachers and previous ones. From a psychometric standpoint it would seem likely that TCE Question 6, which asks students to make a comparison between their current teacher and other teachers they have had, includes a complex set of variables difficult to measure and heavily influenced by each individual’s previous experience. On the contrary, TCE Question 1 which asks students directly for the teacher’s overall teaching effectiveness would appear to be more clearly interpretable for students by isolating concrete variables relative to one individual’s teaching effectiveness rather than taking into account several individual’s teaching effectiveness and then comparing them. The issue of isolating factors specific to effective teaching on this study’s instrumentation will be treated in more depth in the limitations section below.

An additional exploratory component of this study was to investigate the relationship between students’ evaluations on these two TCE questions and their
responses on this study’s target instrumentation. In other words, is there a significant relationship between students’ perceptions regarding what should happen in effective instruction, what is concretely happening, and how these two variables relate to how students evaluate their teachers? When students’ responses on the Effective Teacher questionnaire and the two columns of the Evaluation Questionnaire are included in a multiple regression model with TCE Question 1 (“What is your overall rating of this instructor’s teaching effectiveness?”) as the dependent variable, significant correlations resulted. Similar relationships resulted with the mean of TCE Question 6 (“What is your rating of this instructor compared with other instructors you have had?”) with the only exception being the frequency column of the Evaluation Questionnaire which did not demonstrate a significant correlation with TCE Question 6. From a statistical standpoint, these findings would mean that in the case of TCE Question 1, students’ ideals of what effective teachers should be doing, their perceptions of how often certain behaviors take place, and their evaluations of specific FL teaching practices could predict those same students’ responses on the standard TCE instrument to a degree of accuracy beyond chance. The exploratory and complex nature of these comparisons temper any definitive conclusions that could be drawn from these results. However, it does appear that in this study students’ perceptions of what should be done in the FL classroom and what is being done in the classroom correlated to some extent with students’ ratings of their instructors on two items most related to teaching from the university’s standard evaluation instrument—Teacher Course Evaluation (TCE).
TEACHER COMPARISONS

Another way to approach the analysis of teacher-student comparisons is to focus on individual teachers and the overall correlation between their responses and their students on each questionnaire. This analysis allows for a comparison of questionnaires as well as a more detailed glimpse into possible relationships between teachers’ self-reported demographics and resulting teacher-student correlations on each questionnaire. Of particular interest is the overall number of teachers on each questionnaire whose responses resulted in significant correlations different from zero. On the Effective Teacher Questionnaire where teachers and students reflected on what effective teachers should do ideally in the FL classroom, 17 of 49 teachers had responses across all items which correlated significantly with their students. In the case of the frequency column of the Evaluation Questionnaire where teachers and students reflected on how often participating classroom teachers actually implemented specific practices in their classes, an overwhelming 44 of 49 teachers responded in such a way as to correlate significantly with their students. The lowest number of teachers with significant correlations resulted on the evaluation column of the Evaluation Questionnaire, with a mere 7 teachers responding in a parallel fashion to their students when asked how effective they are in carrying out the target teaching practices.

The fact that participating teachers and students perceived how often teachers integrated target teaching behaviors in a similar fashion appears rather logical, given the somewhat objective nature of the question “How often . . .?” Also quite rational is the finding that teachers’ and students’ perceptions of what effective foreign language
teachers *should* do (Effective Teacher Questionnaire) resulted in only 17 of 49 teachers with significant correlations since each group surely has had diverse experiences in previous FL classes which has led to differing ideals. The role of teacher training might also have exerted an impact on teachers’ ideals and how they coincide or differ with their students’. More intriguing than these two findings is the fact that only seven of the total 49 teachers evaluated themselves in a similar manner as their students across all 21 items on the evaluation column of the Evaluation Questionnaire. As far as correlations of teachers’ self-evaluations and students’ evaluations of them, a large majority of teachers do not appear to perceive their teaching effectiveness on this study’s questionnaire in the same way as their students.

Teacher 25 was the only participating instructor whose responses correlated significantly with students on both the Effective Teacher Questionnaire and both columns of the Evaluation Questionnaire. This teacher was a female, non-native speaker of Arabic with more than ten years of language-related teaching. This instructor also had taken five or more formal teacher training courses and had taught the specific class for which she was responding four or more times. The responses from Teachers 2, 8, 12, 27, and 48 correlated with their students significantly on both columns of the Evaluation Questionnaire, i.e., frequency and evaluation column, but not the Effective Teacher Questionnaire. Teachers 2, 27, and 48 were native speakers of German, Turkish, and Japanese respectively while Teachers 8 and 12 were non-native speakers of Spanish. The only teacher of these five who reported having taken fewer than five formal teacher-training courses was Teacher 12. This same teacher, Teacher 12, did not respond to the
item requesting the number of years experience in teaching language-related classes while the other four reported at least 3-5 years, with two teachers reporting more than 10 years of experience. Of those few teachers who had significant correlations across multiple questionnaires, apparently only Teacher 12 did not have substantial experience teaching nor taking formal teacher-training courses. Table 28 summarizes the Pearson product-moment correlation coefficients for each teacher on the Effective Teacher Questionnaire (Effect.) and both the frequency (Freq.) and evaluation (Eval.) columns of the Evaluation Questionnaire.
Table 28. Summary of Pearson Product-moment Correlations between Teachers and Students by Teacher and Across Both Questionnaires.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Pearson Product-moment Correlations by Questionnaire</th>
<th>Teacher</th>
<th>Pearson Product-moment Correlations by Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.357</td>
<td>.824*</td>
<td>-.030</td>
</tr>
<tr>
<td>2</td>
<td>.276</td>
<td>.600*</td>
<td>.538*</td>
</tr>
<tr>
<td>3</td>
<td>-.006</td>
<td>.527*</td>
<td>.445</td>
</tr>
<tr>
<td>4</td>
<td>.493*</td>
<td>.854*</td>
<td>.415</td>
</tr>
<tr>
<td>5</td>
<td>.463*</td>
<td>.783*</td>
<td>-.142</td>
</tr>
<tr>
<td>6</td>
<td>.121</td>
<td>.724*</td>
<td>.265</td>
</tr>
<tr>
<td>7</td>
<td>.395</td>
<td>.538*</td>
<td>.364</td>
</tr>
<tr>
<td>8</td>
<td>.379</td>
<td>.666*</td>
<td>.604*</td>
</tr>
<tr>
<td>9</td>
<td>.364</td>
<td>.565*</td>
<td>.237</td>
</tr>
<tr>
<td>10</td>
<td>.304</td>
<td>.534*</td>
<td>.131</td>
</tr>
<tr>
<td>11</td>
<td>.642*</td>
<td>.631*</td>
<td>.229</td>
</tr>
<tr>
<td>12</td>
<td>.186</td>
<td>.736*</td>
<td>.485*</td>
</tr>
<tr>
<td>13</td>
<td>.186</td>
<td>.748*</td>
<td>.067</td>
</tr>
<tr>
<td>14</td>
<td>.409*</td>
<td>.504*</td>
<td>.391</td>
</tr>
<tr>
<td>15</td>
<td>.410*</td>
<td>.731*</td>
<td>-.236</td>
</tr>
<tr>
<td>16</td>
<td>.131</td>
<td>.611*</td>
<td>.032</td>
</tr>
<tr>
<td>17</td>
<td>.545*</td>
<td>.330</td>
<td>.189</td>
</tr>
<tr>
<td>18</td>
<td>.437*</td>
<td>.547*</td>
<td>.401</td>
</tr>
<tr>
<td>19</td>
<td>.440*</td>
<td>.817*</td>
<td>.102</td>
</tr>
<tr>
<td>20</td>
<td>.497*</td>
<td>.749*</td>
<td>.365</td>
</tr>
<tr>
<td>21</td>
<td>.015</td>
<td>.219</td>
<td>.596*</td>
</tr>
<tr>
<td>22</td>
<td>.214</td>
<td>.572*</td>
<td>.221</td>
</tr>
<tr>
<td>23</td>
<td>.300</td>
<td>.434*</td>
<td>.426</td>
</tr>
<tr>
<td>24</td>
<td>.134</td>
<td>.708*</td>
<td>.246</td>
</tr>
<tr>
<td>25</td>
<td>.577*</td>
<td>.439*</td>
<td>.724*</td>
</tr>
</tbody>
</table>

*significant at an alpha level of .05.

Interpreting these findings presented in Table 28 accurately may be difficult due to the multiple reasons why teachers differ from students in perceiving what should happen in the FL classroom, what is happening, and how effectively the teacher is making it happen. The only teacher (#25) with statistically significant correlations on all
three measures was an experienced teacher with formal training in language pedagogy. Similarly, the other teachers (#2, #8, #12, #27, and #48) whose responses to both columns of the Evaluation Questionnaire correlated significantly with their students’ appeared to be experienced and well-trained with the only exception being Teacher #12 among the group.

In the case of this study where correlation analyses were used, experienced, well-trained teachers seemed to be in close agreement with their students. Nevertheless, other crucial factors such as personality type might serve to clue in teachers with less experience, should they have very observant, analytical dispositions. The low numbers of teachers whose responses correlated with students also begs the question, “Why so many differences between teachers and students perceptions?” This question becomes more pressing when the training and experience of all teachers is examined more closely. Fifty-eight percent of participating teachers reported having taken 2 or more teacher-training courses and 67% reported having 3 or more years of teaching experience (38 % had more than 6 years experience).

Whether a teacher has years of training and experience or little experience and training, teachers should be careful not to assume that their students are perceiving what should happen or even what is happening in the same way as their teachers. Both experienced and less-experienced teachers may have very different perceptions than their students of what should take place in the classroom and what does take place for a variety of reasons. For example, it seems logical that the well-trained, experienced teacher would have prepared and planned classroom activities based on proven techniques and strategies
that have worked with students over the years. This same sort of teacher may even be well-versed in the latest thinking in FL pedagogy and have attended in-service trainings. In contrast, the less-experienced teacher with little training most likely would be experimenting with different strategies wondering whether they will work or not. It might be that well-trained, experienced teachers do not share with their classes their rationale or objectives behind certain activities or approaches to FL teaching, assuming that the effectiveness is self evident to students. In the case of less-experienced teachers, they might not be sharing their thinking with students for fear of losing face before students or not knowing exactly why they do the things they do, and, therefore, being unable to provide students with coherent, theoretically sound explanations. Suffice it to say, regardless of teachers’ training or experience, teachers must not assume that they are perceiving ideal classroom practices and their own classroom practices in the same way as their students.

LANGUAGE AND LEVEL

The type of language, less commonly taught languages (LCTL) or commonly taught languages (CTL), resulted as a significant main effect in the multiple ANOVA analyses for more items on each questionnaire (Effective Teacher and Evaluation) than level or the interaction between language and level. On the Effective Teacher Questionnaire, only four items demonstrated a statistically significant effect for language with two resulting in higher agreement from LCTL students than CTL students (Item 4—have students use language outside of class; Item 9—be as knowledgeable about culture as language). The remaining two demonstrated a significant effect in the opposite
direction with CTL students expressing greater agreement than LCTL students on Items 2 (assess group tasks) and 10 (not grade production for accuracy).

On the frequency column of the Evaluation Questionnaire, nine items exhibited a statistically significant effect for language with six reflecting greater perceived frequency of teacher integration of target strategies among LCTL classes than CTL classes for the following items: 3 (has students use language outside of class), 4 (corrects oral mistakes directly), 7 (grades for grammatical accuracy), 8 (corrects oral mistakes indirectly), 9 (has students respond to physical commands—TPR), and 14 (presents grammar in real-world context). CTL students perceived their teachers’ integration of the following techniques in the classroom more frequently than did LCTL students: Item 1 (uses computer-based technologies), Item 2 (devotes time for culture), and Item 12 (uses small-group work).

In reflecting on their teachers’ effectiveness in using each technique, language again resulted as the most common statistically significant independent variable. All five items that demonstrated an effect for language resulted in LCTL students perceiving greater effectiveness for their teachers than did CTL students, i.e., Items 3 (has students use language outside of class), 11 (encourages students to speak the FL in class), 14 (presents grammar in real-world context), 19 (engages students in information gap activities), and 21 (uses FL in the classroom).

The rate of significant effects for level in the ANOVA analyses was much lower than for language. Two of the 24 items (Item 8—only correct oral errors indirectly, Item 14—require students to speak FL first day of class) demonstrated significant effects for
level on the Effective Teacher Questionnaire, both of which resulted in greater agreement among second-year students than first-year students. Only one item (Item 10—addresses errors by explaining why) on the frequency column and one on the evaluation column (Item 1—uses computer-based technologies) of the Evaluation Questionnaire resulted statistically significant in the ANOVA models for level with first-year students perceiving greater frequency and effectiveness of these target behaviors by their teachers than second-year students.

Some of these findings relative to the impact of level on students’ responses to the aforementioned items appear quite logical. Second-year students expressed stronger agreement that effective foreign language teachers should require students to speak the FL the first day of class. Students who are in their second year of language study would have had more experience in the FL classroom, and, subsequently, more experience articulating the language, leading to more confidence in speaking it—even from the first day of class. A possible explanation behind second-year students’ preference for indirect oral correction might also relate to increased experience in the classroom. While first-year students may begin their language study by welcoming direct error correction, considering it as the best approach to uprooting and eradicating faulty notions of grammar early on, second-year students may have grown weary of the affective impact of direct oral correction on their confidence. Similarly, first-year students’ preference for explicit explanations following error correction might also be related to their increased desire to directly and immediately deal with errors in their interlanguage in hopes of eliminating them. Finally, first-year students evaluated their teachers’ use of computers
higher than did second-year students and this might again relate to experience with the use of computers in language learning. Foreign language students who have seen the computer used in previous classes, as may be the case for second-year students, most likely will not be as impressed as they were when they first encountered computer-based approaches to language teaching. The case may be that, simply, the novelty of computers has lost its allure for second-year students.

The interaction between language and level only resulted statistically significant on one item across both questionnaires: Item 17 (uses real-life materials in the classroom) from the frequency column of the Evaluation Questionnaire. According to the ANOVA analyses, second-year LCTL students perceived that their teachers used real-life materials more frequently than did first-year LCTL students. On the contrary, second-year CTL students perceived that their teachers used real-life materials less frequently than did their first-year counterparts in CTL classes. Why second-year LCTL students and first-year CTL students perceived that their teachers used real-life materials more than their counterparts in the other level is perplexing? Unless, of course, students’ perceptions are truly reflective of what these participating teachers, in fact, did in their classrooms.

As mentioned previously, any conclusions drawn from these data must be considered tentative and exploratory, given the difficult nature of interpreting such complex results. Nevertheless, the overwhelming proportion of statistically significant effects that resulted for language as compared to level or the interaction between language and level make it apparent that in this study there exist differences in perceptions between LCTL students and CTL students on certain issues fundamental to
FL pedagogy. Some of these may be due to truly differential teaching practices by CTL and LCTL teachers as reflected by students’ responses on the frequency and evaluation columns of the Effective Teacher Questionnaire. However, on the Effective Teacher Questionnaire where both CTL and LCTL students were asked to reflect on what effective foreign language teachers should do in general, it is interesting to note that LCTL students felt more strongly than CTL students that teachers should have students using the FL outside of class. They also felt more strongly about the need for effective foreign language teachers to be just as knowledgeable about the culture as the language. The demographic profile of LCTL students and their motivations for taking LCTL classes needs much more attention in the research in order to better isolate why LCTL students value more than CTL students teachers who require the use of the FL outside of class and who possess as much knowledge about the FL culture as language. It might be that LCTL students are aware of the more limited opportunities for using their languages of study outside of class and the difficulty in finding authentic sources for cultural knowledge, and, therefore, want teachers to find opportunities for them outside of the classroom to use the language and interact with the culture. Moreover, LCTL students could possibly be much more aware than CTL students of the strong cross-cultural and cross-linguistic differences between their languages of study and English.

Furthermore, since Italian was the only LCTL language that was not typologically different from the CTLs, an argument could be made that students and teachers drawn to typologically different languages from the commonly taught languages, Spanish and French (Romance) and German (Germanic), have more disparate notions of appropriate
teaching practices in their classrooms. Due to the unique nature of the LCTLs, it follows that several of the items producing significant main effects among LCTLs related to target language use, error correction, and grammar teaching. Teachers and students in languages such as Japanese, Arabic, Hebrew, Greek, and Turkish would seem to deal with unique issues of target language use, error correction, and grammar teaching, given the more distant relationship between the orthography and grammar of students’ native language, English, and their chosen language of study. Additionally, LCTL students may represent a more academically inclined and intrinsically motivated group of students than the CTL students. It seems less probable that a university student with minimal interest in FL who just wants to get the university’s requirement out of the way would enroll in Arabic, Turkish, or Hebrew rather than French, German, or Spanish. With that said, there may be heritage students who have minimal interest in FL learning and who enroll in Japanese, Hebrew, or Arabic finding it the easiest way to fulfill a FL requirement.

With this specific population, the number of items on these questionnaires that demonstrated statistically significant effects for language far surpassed the number for the other two variables. Nevertheless, crucial differences between first and second-year students resulted on items relative to error correction and target language use. Students’ ideals and subsequent evaluations on different FL teaching strategies may change from their first to second year of FL study. Depending on the direction and degree of those changes, their expectations may be more closely aligned with current thinking in SLA theory and pedagogy or more distant, making even more salient Horwitz’s (1988) claim that beginning level FL students may have unrealistic expectations for FL learning. Most
first-year students should certainly not be considered a blank slate with no preconceived notions of FL teaching. Teachers must help students understand how classroom instruction may change as a result of students improving abilities from first to second-year courses. Likewise, teachers of LCTLs must remember that their students will most certainly bring into the classroom a distinct set of expectations and perceptions on certain teaching techniques different from CTL students. In both cases, teachers must address these differences with their students so that by the end of the semester students who find their skills not at the level they would like, will not target the teaching as the sole culprit for unrealized goals.

DISCUSSION

The teachers in this study appear to value communicative approaches to foreign language pedagogy where information exchange takes precedence over mechanical grammar practice. Teachers also felt more strongly that grammar instruction needed to be set in real-world contexts. What makes this finding worth mentioning is not that these teachers value meaningful information exchange over grammar, but that their respective students do not, or at least appear to not, value it nearly as much. This finding may help prepare teachers to confront a population of students who, for the most part, prefer to have grammar teaching take precedence over communicative exchanges in the FL classroom. This finding corroborates the results from Schulz’s (1996, 2001) studies which found that students’ opinions of grammar teaching appeared more favorable than teachers’. Teachers may need to help students understand several principles of second language acquisition (SLA), e.g., interaction hypothesis, negotiation of meaning, output
hypothesis, by providing explanations and justifications for exercises without a grammar focus or assignments not graded primarily for accuracy, but rather effectiveness of communication. Another possible explanation for students’ perceived interest in grammar teaching might be the disconnect between teaching and testing. These last two issues will be discussed more thoroughly in the Theoretical and Pedagogical Implications section.

Teachers’ and students’ responses correlated more when reflecting on how often certain teaching behaviors took place than how effective teachers were in integrating them. Although teachers and students perceived similar frequencies, they differed in perceiving teacher effectiveness. With respect to teacher-student response patterns, the norm was difference and the exception was consensus. In analyzing teacher-student differences with descriptive statistics overall, it appeared that teachers’ responses approached more of what the field at large might consider appropriate for communicative classrooms. For example, teachers seemed more amenable to meaningful information exchange with less interest in integrating or assessing mechanical grammar practice. Unlike the students in Kern’s (1995a) study whose responses reflected some preference for communicative language teaching, the participants in the present study seemed less amenable to communicative approaches and more in favor of grammar-focused instruction. Likewise, teachers agreed more strongly that having students complete tasks in the language was useful. Teachers seemed more enthusiastic about having students use the foreign language more frequently and earlier on than did their students. Students felt that effective foreign language teachers should correct oral mistakes immediately and teachers were not nearly as convinced—a stance on error correction reflective of
communicative approaches to FL pedagogy. Teachers also reported that they used indirect error correction more than direct error correction in their classrooms while students did not concur. Similar to Reber (2001) who found disagreement among participants on controversial issues such as grammar teaching, assessment practices, and target language use, this study exposed differences between students and teachers on the very same topics.

Both the frequency with which teachers perceived themselves to be integrating certain skills and their effectiveness in doing so paralleled what is currently termed ‘communicative’ in the pedagogical literature. In other words, teachers’ opinions of what should be happening in the classroom approached an ‘ideal’ communicative classroom, where students communicate about meaningful topics, complete real-world tasks, use computer-based technologies, engage with the language outside of class, gain exposure to the culture in class, and work in groups or pairs. Unfortunately, in many cases participating teachers’ perceptions of this ‘ideal’ communicative classroom did not parallel their students’. Though teachers’ and students’ perceptions of what was actually happening in the classroom coincided slightly more than they did when reflecting on ideal practices, differences abounded. Teachers’ perceived themselves as integrating the target behaviors more than students and self-evaluated their performance on these same practices as more effective than did students. Apparently, students either have not been formally exposed to current trends in FL pedagogy, or, if they have been exposed, they do not agree with many of its fundamental tenets. In reality, teacher perceptions may be more closely related to what they think they should do based on their training than what
they actually do in the classroom. If, indeed, teachers believe their performance approaches the ideal communicative classroom it may come as a shock when observers, students and supervisors, give them poor evaluations.

Analysis of the relationship between teacher demographics and the degree of teacher-student differences seemed to favor experienced, well-trained teachers in predicting which teachers’ perceptions might be more in sync with students’. Many of the teachers in this sample whose responses correlated significantly with their students’ across questionnaires had language teaching experience and formal, pedagogical training.

Students’ and teachers’ perceptions of effective teaching and their concrete evaluations of teaching clearly represent a complex interplay of numerous factors, interpersonal and pedagogical, as well as social and psychological. Rather than attempting to discover an authoritative, indisputable definition of effective L2 teaching, this study makes headway into providing a profile of how L2 students and their teachers define it, and how their definitions compare. The match, or mismatch, of L2 teachers’ and students’ perceptions of effective teaching in general, coupled with a comparison of their course-specific ratings can create greater understanding of each group’s perspective on effective teaching in the L2 classroom. Not only can teachers and students benefit from this increased awareness, so too can basic language directors, administrators, and teacher trainers. When a teaching supervisor or director is confronted with discontent among teachers or students, it may be helpful to look at each group’s expectations and perceptions of what goes on in the classroom. Furthermore, the results of this study will add to and supplement previous research from which SLA scholars may continue to
reconcile their concrete, empirical findings with perceptions of those most directly involved with L2 learning: teachers and students. If students feel their teachers better understand their perspectives and, likewise, if teachers feel that students better understand theirs, an increase in motivation and satisfaction may result for both groups. Teachers who take the time to assess their students perspectives on FL teaching, either through questionnaires or in-class discussion, will validate their students’ opinions while producing students who feel that their perceptions are an essential consideration in preparing classroom activities.

THEORETICAL AND PEDAGOGICAL IMPLICATIONS

The small number of statistically significant correlations along with frequent averaged teacher-student differences approaching .50 on a 4-point scale leads to the most obvious implication for classroom teaching: Any given group of FL students will most assuredly have differing opinions from their teachers on some salient issue in FL pedagogy. Teachers must not assume that students entering their FL classes share the same perspectives as they do on what it means to learn a language, how to go about it, and which are the most effective means of doing so. Additionally, teachers must not assume that students will accurately perceive how or how often certain techniques are implemented. Teachers may feel that they are doing something frequently when their students may feel that it is done quite infrequently. The same principle applies to a teacher’s perceived effectiveness and how that perception correlates with students’ perceptions. A teacher may find that after spending an entire four-month semester using a
favorite grammar exercise, that students express disfavor and disillusionment on end-of-semester evaluations.

At its core, the results of this study imply that teachers must communicate more with their students regarding their convictions of FL pedagogy and the concrete teaching taking place in the classroom. This study brings to the forefront faulty assumptions some teachers may make about their students, namely, 1) students trust that their teacher will choose those activities that will most spur language learning and are most appropriate, 2) students agree with the teacher’s pedagogical decisions in the classroom and will not question them, 3) students do not have sufficient capacity to understand basic SLA theory and how that can translate into practice.

Short of making each basic language class a methods course, beginning-level instructors would do well to take Horwitz’s (1988) suggestion to assess their students’ perspectives on language learning early in the semester. Taking her recommendation a step further, it might prove useful to not only inquire about students’ ideas concerning language learning, but to preview for students the sorts of concrete pedagogical activities that they will integrate in the classroom. What this study has shown is that it is not only useful to know what students feel about abstract SLA principles as Horwitz (1988) and Kern (1995a) have demonstrated, but also what students’ perceptions are of concrete pedagogical practices in the FL classroom. In fact, teachers might engage students in brief cursory discussions of SLA by first explaining their rationale behind certain activities. After informing students that group and pair work will be used frequently in the classroom, the teacher might help students understand the importance of input,
output, and negotiation of meaning, ensuring to adapt the discussion and terminology to students’ level of understanding.

Three specific areas of FL teaching that teachers must address are error correction, grammar teaching, and group/pair work. Teacher-student mismatches on these three issues may reflect fundamental, theoretical differences between teachers and students in perceiving the whole process of second language acquisition as well as the nature of achievable goals for FL learning. Students who value an emphasis on grammar teaching and overt error correction may feel that second language acquisition is strictly about obtaining declarative knowledge regarding a language’s grammar. This perspective will most likely lead to frustration when students struggle to use the language spontaneously in either spoken or written modalities. Teachers who value accuracy in production but choose to adopt less overt or obtrusive strategies in teaching grammar may end up with disillusioned students whose unrealistic expectations are not met. Furthermore, teachers appear to find more value in group/pair work than their students and feel that they use it more effectively than their students feel they do. Teacher-student disagreement with group/pair work, such a fundamental and frequently used pedagogical tool in these FL classes and many others, warrants discussion in the classroom so that teachers can justify its use and so that students can have their concerns addressed.

Teachers surely do not have the time nor energy to provide a thorough theoretical justification for each and every activity they have students complete. Similarly, teachers need not feel obligated to correct all of their students’ misconceptions regarding SLA and language teaching in general. Nonetheless, students’ perceptions of effective teaching
practices and teacher effectiveness might be altered were teachers to provide a short rationale for selected activities. For example, when engaging students in group work, it may increase student motivation if teachers simply asked students why the activity is being conducted in groups or pairs instead of as an entire class. Some students may pick up on the need to maximize learners’ opportunities for production through small group work but others may not. In summary, this research makes evident the need for teachers to 1) assess students’ perspectives on concrete pedagogical practices in the classroom and 2) selectively share their rationale and justifications for integrating certain activities so as to further inform and motivate their students.

LIMITATIONS OF THIS STUDY

As is the case with most studies, limitations exist which restrict the generalizability of the study’s results. The principal limitation of this study relates to the instrumentation, while the participant sampling procedures present a secondary concern. As Brown (2001) has pointed out, effective item writing can be extremely difficult and one of the most arduous aspects of questionnaire design. This study’s main objective was to make specific, direct comparisons between teachers’ and students’ perceptions of a subset of salient considerations and practices in foreign language teaching. Closed-response, Likert-scale items were used to allow teachers and their students to reflect on what they feel should be happening in FL classrooms, what is actually happening in their classrooms, and how effective their teachers are in implementing each strategy.

Although the items on each questionnaire were carefully worded, piloted with intact classes, and revised several times to ensure intelligibility, some of the items may
have been too specific, vague, or difficult for individual participants to provide an accurate representation of their perspectives. The terminology on the 4-point Likert-scales used on the Effective Questionnaire and the Evaluation Questionnaire might have presented intervals that were not equidistant psychologically to students, and not anchored by direct opposites as required by a true Likert scale. For example, the extreme negative anchor represented by ‘Never’ would be a complete opposite of ‘Always’ instead of ‘Frequently’. Likewise, the evaluation scale used a 4-point scale including ‘Very Effective’ ‘Effective’ ‘Limited Effectiveness’ and ‘Ineffective’ where a more psychometrically sound scale might include the descriptors ‘Very Effective’, ‘Effective’, ‘Ineffective’, ‘Very Ineffective’ so as to create equidistant intervals in the 4-point scale from a semantic and psychological perspective. Although an effort was made to avoid ‘double-barreled’ items where possible, i.e., items with multiple clauses or phrases joined by the conjunctions ‘and’ or ‘or’, space and time constraints led to the inclusion of a limited number of double-barreled items. One or both of these explanations may have contributed to the unclear factor loadings that resulted from the factor analysis. The factor analysis that was conducted produced seven factors, some of which loaded on seemingly unrelated items. From a strictly mathematical and statistical perspective, the factor analysis failed to present a clear picture of how all the items correlated and whether they were truly measuring the target constructs.

More rigorous tests of reliability and factor loadings and subsequent revisions to the items could have proved helpful in making each item more clearly representative of unique factors in foreign language teaching. Nevertheless, the present instrumentation
does not attempt to include all factors deemed essential to effective foreign language
teaching as this study’s primary goal was not to argue for a finite, linear relationship
between the items on the questionnaire and all concrete behaviors empirically proven to
reflect effective foreign language teaching. Instead, the objective of this research was to
justify the inclusion of a subset of relevant considerations in FL pedagogy and provide a
detailed comparison between teachers’ and students’ perspectives of those specific issues.

One delimitation of the study that could be construed as a limitation is the use of a
quantitative, closed-response questionnaire where participants are forced to respond to
only the set of items included on the questionnaire. Participants may have had in mind
several teaching strategies or behaviors characteristic of effective foreign language
teachers that did not appear on the closed-response questionnaires. Also, the exclusion of
qualitative data, i.e., short-answer responses, in the study may be considered by some as a
limitation. Within the research framework adopted for this study, time and human
constraints did not make a qualitative study possible.

Another limitation concerning the instrumentation used in this study relates to the
complex format and structure of the Evaluation Questionnaire. In an attempt to not
conflate students’ and teachers’ perceptions of the frequency of a teaching behavior with
the effective implementation of a teaching behavior, two distinct columns were provided
as explained in Chapter 3. As part of the oral and written instructions, participants were
instructed to skip the evaluation column if they felt that they “Never” observed a certain
behavior from their teachers, since the evaluation column was to reflect how effective
students felt their teachers were in integrating those specific teaching practices. As some
students filled out the evaluation column after stating that they had never observed their teachers perform certain strategies, it was obvious that certain students, and possibly teachers, were reflecting on whether they considered the teaching tactic effective in general—the perspective elicited in the Effective Teacher Questionnaire administered at the beginning of the study. Additionally, on certain items some students and teachers might have struggled to thoughtfully and accurately distinguish between how often teachers performed certain techniques versus how effective they were in executing them. It takes reflective thought to differentiate between, for example, how often a teacher makes “use of English in the foreign language classroom” and how effective they make “use of English in the foreign language classroom.”

A final limitation relates specifically to the Evaluation Questionnaire and the prompt statement used for teachers and students. Differing prompts for teachers—“As a foreign language teaching in my current classes of the same level, I . . .”, and students—“In this class my foreign language teacher . . .”, on the frequency and evaluation columns of the Evaluation Questionnaire prevented the use of inferential statistics to directly compare teachers’ and students’ means so as to rule out differences due to chance.

Finally, teachers were recruited voluntarily with their students and, hence, might not be a representative sample of 1st and 2nd-year language teachers at the University of Arizona. Thus, their interest and knowledge of foreign language teaching methods might be inflated and lead to response patterns and teacher-student differences not necessarily reflective of the average foreign language class. Clearly, the results from this study can only be taken to be representative of the participating population and to a slightly lesser
degree representative of the larger population of foreign language students and teachers at
the University of Arizona. However, the large number of participating students may
warrant tentative comparisons to other post-secondary foreign language settings within
large, public universities.

RECOMMENDATIONS FOR FUTURE RESEARCH

The results of this study point to several issues that future research needs to
address within the field of FL students’ and teachers’ beliefs and perceptions. More
research must explore how and where students formulate their ideas of effective and
ineffective foreign language learning and teaching. Conventional wisdom has been to
look back at learners’ previous foreign language experiences and analyze the impact of
previous teachers and previous experiences on learners’ current perspectives. Rather than
looking solely at previous experiences in FL classes, the concept of ‘apprenticeship of
observation’ needs to be expanded to include students’ experiences in other classes of
other disciplines. Students’ apparent penchant for discrete-point grammar may be a
product of teaching and assessment practices in FL classes or other disciplines where
multiple-choice, true-false, and matching formats are used rather than open-ended
responses evaluated by analytic scales with descriptors. Students may feel more
comfortable with traditional, discrete-point assessment instruments than with descriptive,
analytic rubrics reflective of communicative language teaching because that is how they
are evaluated in other classes. Metaphorically speaking, researchers need to not only look
back at students’ previous experiences in the FL classroom but also look to the side in
contemplating prevalent pedagogies and assessment practices in other disciplines and
how they might influence students’ perceptions of what shape foreign language teaching should take. If in other classes students are repeatedly exposed to assessment instruments and practices which are based on the strict right-wrong dichotomy, they may value that approach in their language learning. Of the numerous facets of FL teaching and learning, grammar appears to provide the best fit for ‘right-wrong’ testing, and many students seem to feel more comfortable knowing that there is one ‘right’ answer rather than a graded continuum of more accurate to less accurate.

As mentioned previously, Horwitz’s (1988) recommendation to assess students’ perspectives on foreign language learning at the beginning of the semester might be supplemented by a questionnaire to determine students’ perspectives of concrete exercises and teaching behaviors that they might encounter later on. Future research needs to explore the cognitive and affective impact on students when teachers share their rationale behind their teaching practices and require students to articulate justifications for certain activities from an SLA perspective. Do teachers who follow this practice produce students who learn more while maintaining greater motivation? Do classes where students’ perspectives on relevant FL teaching practices correlate closely with their teachers’ learn more with greater levels of satisfaction than those students whose perspectives do not correlate closely with their teachers’? This research area might parallel the learning strategy research of the 90’s where students’ and teachers’ general learning and teaching styles were matched in an attempt to determine effects on concrete learning gains.
In reference to students’ and teachers’ evaluations of teaching, more work must be done to expand on the current study’s attempt to determine the relationship between teacher-student differences and students’ summative ratings of teachers. More sophisticated research models with well-designed instruments might allow researchers to make more specific, direct correlations between teacher-student perceptual differences and student ratings. Teachers would be more prone to give heed to differences in teacher-student perceptions were it shown that these differences were accurate predictors of students’ ratings. That is not to say that teachers should only teach using strategies and techniques that students approve of, but that they should try and bridge the gap between their perceptions of effective teaching and their students’.

CONCLUSION

Although a formidable and seemingly endless task, researchers need to continue assessing teachers’ and students’ perceptions of FL teaching, as the field will most assuredly change over time and idiosyncratic perceptions of FL teaching among teachers and students will most definitely remain a reality of the FL classroom. The most practical and far-reaching impact of future research in this area will take place in individual FL classrooms where teachers become researchers who desire to better understand their own perspectives on FL teaching, their students’ perspectives, and how to reconcile the two.
APPENDIX A: Effective Teacher Questionnaire

What are the 4 digits of your birth month and day (mm/dd) and the first 3 letters of your mother's first name, e.g. 09-18-JAN?
0123 4567 890123 4567 890123 4567 890123 4567 890123 4567 890123 4567 890123 4567 890123 4567 89

The Effective Foreign Language Teacher

Instructions: Please reflect on your personal beliefs regarding what characterizes effective foreign language teaching. Carefully read each statement and indicate to what extent you agree or disagree by bubbling in the acronym that best describes your opinion. There are no right or wrong answers, just those that are right for you. Your sincere, personal responses will guarantee the success of the study. Thank You!

Effective foreign language teachers should...

1. Frequently use computer-based technologies (internet, CD-ROM, email) in teaching the language.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

2. Base at least some part of students' grades on completion of assigned group tasks.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

3. Devote as much time as to the teaching of culture as to the teaching of language.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

4. Require students to use the language outside of class with other speakers of the language (e.g., internet, email, clubs, community events, etc.).
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

5. Not correct students immediately after they make a mistake in speaking.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

6. Allow students to respond to test questions in listening and reading via English rather than the foreign language.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

8. Only correct students indirectly when the produce oral errors instead of directly, e.g., correctly repeating back to them rather than directly stating that they are incorrect.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

9. Be as knowledgeable about the culture(s) of those who speak the language as the language itself.
   - Strongly Agree
   - Agree
   - Disagree
   - Strongly Disagree

10. Not grade language production (i.e., speaking and writing) primarily for grammatical accuracy.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

11. Teach the language primarily by having students complete specific tasks (e.g., finding out prices of rooms and rates at a hotel) rather than grammar-focused exercises.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

12. Have students respond to commands physically in the foreign language (e.g., "stand up," "pick up your book," etc.).
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

13. Address errors by immediately providing explanations as to why students' responses are incorrect.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

14. Require students to speak in the foreign language beginning the first day of class.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

15. Not use predominantly small groups or pair work to complete activities in class.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

16. Mostly use activities that practice specific grammar points rather than activities whose goal is to merely exchange information.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

17. Ask students to begin speaking the foreign language only when they feel they are ready to.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

18. Not present a particular grammar point without illustrating how the structure is used in a specific, real-world context.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

19. Speak the foreign language with native-like control of both grammar and accent.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

20. Teach grammar by giving examples of grammatical structures before explaining the grammar rules.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

21. Use predominantly real-life materials (e.g., music, pictures, foods, clothing) in teaching both the language and the culture rather than the textbook.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

22. Not simplify or alter how they speak so that students can understand every word being said.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

23. Base at least some part of students' grades on their ability to interact with classmates successfully in the foreign language.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree

24. Use activities where students have to find out unknown information from classmates using the foreign language.
    - Strongly Agree
    - Agree
    - Disagree
    - Strongly Disagree
APPENDIX B: Student Information Questionnaire

What are the 4 digits of your birth month and day (mm/dd) and the first 3 letters of your mother's first name, e.g. 0916-JAN?

What is your class number?

Student Information

Personal Data:
1-What is your gender?
- Male
- Female

2-What is your age?
- 19-20
- 21-23
- 24-27
- 28-35
- 36-50
- 50+

Current Academic Data:
3-Which of the following best describes your current status as a student at the university?
- Freshman
- Sophomore
- Junior
- Senior
- Graduate
- Non-degree

4-What is your approximate, overall GPA at the university?
- 4.0-3.5
- 3.4-3.0
- 2.9-2.5
- 2.4-2.0
- 1.9-0.0
- N/A

5-Why are you taking this class?
- Major
- Minor
- Foreign language requirement
- Personal interest

6-How difficult do you find this specific foreign language class to be?
- Very easy
- Easy
- Average
- Difficult
- Very difficult

7-Truthfully, what grade do you expect from this particular class?
- A
- B
- C
- D
- F
- N/A

Foreign/Second Language
8-Did you hear the language of this class frequently as a child?
- Yes
- No

9-Did you speak the language of this class frequently as a child?
- Yes
- No

10-As children, did your parents or grandparents speak the language of this class?
- Yes
- No

11-How old were you when you started learning English?
- English is my first/native tongue
- Between 1-10 yrs. old
- 11-20 yrs. old
- After 20

12-Which best describes your self-assessed language proficiency in the language of this class?
- Beginner
- Intermediate
- Advanced
- Near-native

13-How much formal education have you had in the language of this class? (please check all that apply)
- Junior high/High school:
- 1 year
- 2 years
- 3-4 years
- More than 4 years

- Community College/University:
- 1-2 semesters
- 3-4 semesters
- More than 4 semesters

14-How long have you lived abroad consecutively in a country where the language of this class is spoken?
- Never
- 1-3 mos.
- 1-6 mos.
- 7 mos-2 years.
- More than 2 years

15-Have you studied another language other than this language and your native language?
- Yes
- No

If yes, how long?
- 1 semester
- 1 year
- 2-3 years
- 4+

16-How would you describe your overall experience with classroom foreign language learning?
- Unsatisfactory
- Poor
- Neutral
- Good
- Excellent
## APPENDIX C: Summary of Students’ Responses to Student Information Questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Data</th>
<th>Question</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Male</td>
<td>671 (45%)</td>
<td>Q9</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>831 (55%)</td>
<td>No</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-20</td>
<td>1080 (72%)</td>
<td>Q10</td>
<td>Yes</td>
</tr>
<tr>
<td>21-23</td>
<td>271 (18%)</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>24-27</td>
<td>78 (5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-35</td>
<td>40 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-50</td>
<td>25 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 +</td>
<td>8 (.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>614 (41%)</td>
<td>Q11</td>
<td></td>
</tr>
<tr>
<td>Sophomore</td>
<td>406 (27%)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>243 (16%)</td>
<td>85 (6%)</td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td>176 (12%)</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>49 (3%)</td>
<td>1417 (94%)</td>
<td></td>
</tr>
<tr>
<td>Non-degree</td>
<td>14 (.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0 - 3.5</td>
<td>465 (31%)</td>
<td>Q12</td>
<td>Beginner</td>
</tr>
<tr>
<td>3.4 - 3.0</td>
<td>472 (31%)</td>
<td>Intermediate</td>
<td></td>
</tr>
<tr>
<td>2.9 - 2.5</td>
<td>337 (22%)</td>
<td>Advanced</td>
<td></td>
</tr>
<tr>
<td>2.4 - 2.0</td>
<td>146 (10%)</td>
<td>Near-native</td>
<td></td>
</tr>
<tr>
<td>1.9 - 0</td>
<td>40 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>41 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>major</td>
<td>97 (7%)</td>
<td>Q13</td>
<td></td>
</tr>
<tr>
<td>minor</td>
<td>234 (16%)</td>
<td>Jr. High/High School</td>
<td></td>
</tr>
<tr>
<td>FL requir.</td>
<td>865 (58%)</td>
<td>1 year</td>
<td></td>
</tr>
<tr>
<td>personal int.</td>
<td>245 (16%)</td>
<td>2 years</td>
<td></td>
</tr>
<tr>
<td>comb. of 2 or more</td>
<td>61 (4%)</td>
<td>3-4 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 4 yrs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comm. Coll./Univers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-2 sem.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3-4 sem.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 4 sem.</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very easy</td>
<td>41 (3%)</td>
<td>Q14</td>
<td>Never</td>
</tr>
<tr>
<td>Easy</td>
<td>186 (12%)</td>
<td>1-3 wks.</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>771 (51%)</td>
<td>1-6 mos.</td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>434 (29%)</td>
<td>7mos. - 2 yrs.</td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td>70 (5%)</td>
<td>More than 2 yrs.</td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>657 (44%)</td>
<td>Q15</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>647 (43%)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>168 (11%)</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>10 (.7%)</td>
<td>If yes, how long?</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>1 semester</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 year</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>840 (56%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>139 (9%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3 yrs.</td>
<td>4+</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>14 (.9%)</td>
<td>300 (20%)</td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td>300 (20%)</td>
<td>223 (15%)</td>
<td></td>
</tr>
<tr>
<td>Q16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>269 (18%)</td>
<td>43 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1233 (82%)</td>
<td>89 (6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>436 (29%)</td>
<td>745 (50%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>189 (13%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D: Teacher Information Questionnaire

What are the 4 digits of your birth month and day and the first 3 letters of your mother’s first name, e.g. 0916-JAN?

<table>
<thead>
<tr>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
</tr>
</thead>
</table>

What is your class number?

<table>
<thead>
<tr>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
<th>0123456789</th>
</tr>
</thead>
</table>

**Teacher Information**

**Personal Data:**

1. What is your gender? [ ] Male  [ ] Female
2. What is your age? 21-30  31-40  41-50  50+
3. As a child, did you frequently hear the language you teach?
   [ ] Yes  [ ] No
4. As a child, did you frequently speak the language you teach?
   [ ] Yes  [ ] No
5. As children, did your parents or grandparents hear and speak the language you teach?
   [ ] Yes  [ ] No
6. Which term best defines your language proficiency in the language you teach?
   [ ] Adequate  [ ] Good  [ ] Excellent  [ ] Near-native  [ ] Native
7. Which term best defines your familiarity with the culture(s) affiliated with the languages you teach?
   [ ] Adequate  [ ] Good  [ ] Excellent  [ ] Near-native  [ ] Native
8. How long have you lived abroad consecutively in a country where the language of this class is spoken?
   never  1-3 mos.  3-6 mos.  6 mos.-1 year  More than a year  native speaker
9. If you are not from the U.S., how long have you resided in the U.S.?
   less than 4 mos.  4 mos.-1 year  1-2 years  2+ years

**Current Academic Status:**

10. Which of the following best describes your current status at the university?
    [ ] Grad student in same language dept.  [ ] Grad student in another dept.
    [ ] Adjunct faculty  [ ] Non-student instructor  [ ] Full-time lecturer  [ ] Assistant, associate or full professor

**Training:**

11. How many teacher training courses have you formally taken (e.g. teaching methods, testing)?
    [ ] None  1  2-4  5+  

**Previous Language Teaching Experience:**

12. Overall, how many years have you been teaching language-related courses?
    [ ] Less than 1 yr.  1-2 yrs.  3-6 yrs.  6-10 yrs.  More than 10 yrs.
13. How many years have you been teaching language-related courses in the United States?
    [ ] Less than 1 yr.  1-2 yrs.  3-6 yrs.  6-10 yrs.  More than 10 yrs.

Which of the following levels and course types have you taught? (please check all that apply)

14. General Level: (check all that apply)
    [ ] Elementary/Junior or high school  [ ] Undergraduate university courses  [ ] Community college
    [ ] Graduate courses
15. Course Level: (check all that apply)
    [ ] 1st-2nd year courses, i.e. 100–200 level  [ ] 3rd-4th year courses, i.e. 300–400 level
16. Course Type: (check all that apply)
    [ ] Basic 4 skills: listening  [ ] Literature  [ ] Composition  [ ] Grammar  [ ] Other

**Current Teaching Responsibilities:**

17. Have you taught the class you are currently teaching before?
    [ ] Yes  [ ] No
   If yes, how many times?
   [ ] 1  [ ] 2-3  [ ] 4+  

APPENDIX E: Summary of Teachers’ Responses to Teacher Information Questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Data</th>
<th>Question</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td></td>
<td>Q10</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>12 (25%)</td>
<td>grad. stud. in same dept.</td>
<td>22 (45%)</td>
</tr>
<tr>
<td>Female</td>
<td>37 (76%)</td>
<td>grad. stud. in another dept.</td>
<td>16 (33%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>adj. fac., non-stud. instr.</td>
<td>5 (10%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>full-time lecturer</td>
<td>4 (8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>assist., assoc., or full prof.</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td>Q11</td>
<td></td>
</tr>
<tr>
<td>21-25</td>
<td>13 (27%)</td>
<td>none</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>26-30</td>
<td>14 (29%)</td>
<td>1</td>
<td>20 (41%)</td>
</tr>
<tr>
<td>31-35</td>
<td>10 (21%)</td>
<td>2-4</td>
<td>13 (27%)</td>
</tr>
<tr>
<td>36-40</td>
<td>3 (6%)</td>
<td>5+</td>
<td>15 (31%)</td>
</tr>
<tr>
<td>40+</td>
<td>8 (17%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td>Q12</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>30 (61%)</td>
<td>less than 1 year</td>
<td>7 (15%)</td>
</tr>
<tr>
<td>No</td>
<td>19 (39%)</td>
<td>1-2 yrs.</td>
<td>9 (19%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3-5 yrs.</td>
<td>14 (29%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-10 yrs.</td>
<td>7 (15%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 10 yrs.</td>
<td>11 (23%)</td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td>Q13</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>30 (61%)</td>
<td>less than 1 year</td>
<td>12 (25%)</td>
</tr>
<tr>
<td>No</td>
<td>19 (39%)</td>
<td>1-2 yrs.</td>
<td>11 (23%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3-5 yrs.</td>
<td>13 (27%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-10 yrs.</td>
<td>7 (14%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>more than 10 yrs.</td>
<td>6 (12%)</td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td>Q14</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>29 (59%)</td>
<td>elem./jun. or high school</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>20 (41%)</td>
<td>undergrad. univ.</td>
<td>28 (57%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comm. college</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>grad. courses</td>
<td>1 (2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comb. of 2 or more</td>
<td>20 (41%)</td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td>Q15</td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>0</td>
<td>1st-2nd yr. courses</td>
<td>37 (76%)</td>
</tr>
<tr>
<td>Good</td>
<td>4 (8%)</td>
<td>3rd-4th yr. courses</td>
<td>12 (25%)</td>
</tr>
<tr>
<td>Excellent</td>
<td>11 (23%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near-native</td>
<td>9 (18%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native</td>
<td>25 (51%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td>Q16</td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>0</td>
<td>basic 4 skills/culture</td>
<td>23 (47%)</td>
</tr>
<tr>
<td>Good</td>
<td>8 (16%)</td>
<td>literature</td>
<td>0</td>
</tr>
<tr>
<td>Excellent</td>
<td>14 (29%)</td>
<td>composition</td>
<td>0</td>
</tr>
<tr>
<td>Near-native</td>
<td>3 (6%)</td>
<td>grammar</td>
<td>0</td>
</tr>
<tr>
<td>Native</td>
<td>24 (49%)</td>
<td>other</td>
<td>3 (6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comb. of 2 or more</td>
<td>23 (47%)</td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td>Q17</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>4 (8%)</td>
<td>Yes</td>
<td>35 (71%)</td>
</tr>
<tr>
<td>1-3 wks.</td>
<td>0</td>
<td>No</td>
<td>14 (29%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-6 mos.</td>
<td>7mos. – 1 yr.</td>
<td>More than a year</td>
<td>Native speaker</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 (18%)</td>
<td>3 (6%)</td>
<td>11 (23%)</td>
<td>22 (45%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q9:
- less than 4 mos.
- 4 mos. – 1 year
- 1-2 years
- 2+ years

1 (4%)
2 (7%)
2 (7%)
23 (82%)
APPENDIX F: Evaluation Questionnaire (Student Version)

What are the 4 digits of your birth month and day and the first 3 letters of your mother's first name, e.g. 0916-JAN?

1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7
1-7 1-7 1-7 1-7 1-7 1-7 1-7 1-7

Language Teaching Observation & Evaluation

Instructions: In reflecting on your current foreign language teacher and the teaching performed in your class during this semester, please carefully indicate: 1) how often your teacher performs each behavior in your class and 2) when the behavior does take place, how effectively your teacher performs it to aid in your personal language learning. If you “Never” observe a certain behavior then please mark “NA.” Your sincere, personal responses will guarantee the success of the study, therefore, there are no right or wrong answers. Thank You!

<table>
<thead>
<tr>
<th>How often?</th>
<th>How effective?</th>
</tr>
</thead>
<tbody>
<tr>
<td>F=Frequently</td>
<td>VE=Very effective</td>
</tr>
<tr>
<td>S=Sometimes</td>
<td>E=Effective</td>
</tr>
<tr>
<td>R=Rarely</td>
<td>LE=Limited effectiveness</td>
</tr>
<tr>
<td>N=Never</td>
<td>IE=Ineffective</td>
</tr>
<tr>
<td>NA=Not applicable</td>
<td></td>
</tr>
</tbody>
</table>

In this class, my foreign language teacher...

1-uses computer-based technologies other than word processing (e.g., internet, CD-ROM, email) in teaching the language.

2-devotes time to the teaching of culture(s) of those who speak the foreign language.

3-provides opportunities for students to use the language outside of class with other speakers of the language (e.g., internet, email, clubs, community events, etc.).

4-corrects students directly when they make oral mistakes (e.g., by specifically stating that a certain structure they have used is incorrect).

5-makes use of English in the foreign language classroom.

6-has students complete tasks with a real-life purpose in the foreign language (e.g., planning a vacation using the internet or travel brochures).

7-grades exercises in the foreign language (i.e., speaking, writing) for grammatical accuracy.

8-correction students indirectly when they make oral errors (e.g., by correctly repeating back to them, pausing after the error, asking them to repeat what they said, etc.).

9-has students respond to commands physically in the foreign language (e.g., “stand up,” “pick up your book,” etc.).

10-addresses errors by providing explanations as to why students’ language is incorrect.

11-encourages students to speak the foreign language within the classroom when others.

12-uses small-group work to complete in-class activities in the foreign language.

13-uses activities whose goal is to exchange information rather than practicing specific grammar points (e.g., talking about an individual’s daily activities vs. conjugating verbs in table form).

14-presents lessons that emphasize particular grammar aspects by illustrating how the structure is used in a real-world context (e.g., the future tense in setting goals or New Year’s resolutions).

15-demonstrates native-like control in expressing himself/herself in the foreign language (i.e., uses accurate grammar and native-like accent).

16-teaches grammar by explaining the grammatical rules before giving concrete examples for students to consider.

17-uses real-life materials (e.g., music, pictures, food, clothing) in the foreign language classroom.

18-alters how she speaks so that students can understand what is being said.

19-uses activities where students have to find out information from classmates using the foreign language.

20-grades activities where students must interact with classmates in the foreign language (e.g., in pairs or small groups to complete role plays, dialogues, presentations, etc.).

21-uses the foreign language in the classroom.
APPENDIX G: Evaluation Questionnaire (Teacher Version)

What are the 4 digits of your birth month and day and the first 3 letters of your mother's first name, e.g. 0918-JAN?

Language Teaching Observation & Evaluation

Instructions: In reflecting on your current teaching in your class(es) of the same level during this semester, please carefully indicate: 1) how often you perform each teaching behavior in your class and 2) when you do perform a behavior, how effectively you perform it to aid your students in their personal language learning. If you "Never" perform a certain behavior then please mark "NA." Your sincere, personal responses will guarantee the success of the study, therefore, there are no right or wrong answers. Thank You!

<table>
<thead>
<tr>
<th>How often?</th>
<th>How effective?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent</td>
<td>Very Effective</td>
</tr>
<tr>
<td>Sometimes</td>
<td>Effective</td>
</tr>
<tr>
<td>Rarely</td>
<td>Limited Effectiveness</td>
</tr>
<tr>
<td>Never</td>
<td>Ineffective</td>
</tr>
<tr>
<td>NA</td>
<td>Not Applicable</td>
</tr>
</tbody>
</table>

As a foreign language teacher in my current class(es) of the same level, I...

1-use computer-based technologies other than word processing (e.g., internet, CD-ROM, email) in teaching the language.

2-devote time to the teaching of culture(s) of those who speak the foreign language.

3-provide opportunities for students to use the language outside of classes with other speakers of the language (e.g., internet, email, clubs, community events, etc.).

4-correct students directly when they make oral mistakes (e.g., by specifically stating that a certain structure they have used is incorrect).

5-make use of English in the foreign language classroom.

6-have students complete tasks with a real-life purpose in the foreign language (e.g., planning a vacation using the internet or travel brochures).

7-grade exercises in the foreign language (i.e., speaking, writing) for grammatical accuracy.

8-correct students indirectly when they make oral errors (e.g., by correctly repeating back to them, pausing after the error, asking them to repeat what they said, etc.).

9-have students respond to commands physically in the foreign language (e.g., "stand up," "pick up your book," etc.).

10-address errors by providing explanations as to why students' language is incorrect.

11-encourage students to speak the foreign language within the classroom with others.

12-use small-group work to complete in-class activities in the foreign language.

13-use activities whose goal is to exchange information rather than practicing specific grammar points (e.g., talking about an individual's daily activities vs. conjugating verbs in table form).

14-present lessons that emphasize particular grammar aspects by illustrating how the structure is used in a real-world context (e.g., the future tense in setting goals or New Year's resolutions).

15-demonstrate native-like control in expressing myself in the foreign language (i.e., use accurate grammar and native-like accent).

16-teach grammar by explaining the grammatical rules before giving concrete examples for students to consider.

17-use real-life materials (e.g., music, pictures, food, clothing) in the foreign language classroom.

18-alter how I speak so that students can understand what is being said.

19-use activities where students have to find out information from classmates using the foreign language.

20-grade activities where students must interact with classmates in the foreign language (e.g., in pairs or small groups to complete role plays, dialogues, presentations, etc.).

21-use the foreign language in the classroom.
APPENDIX H: TCE Questions

What are the 4 digits of your birth month and day and the first 3 letters of your mother’s first name, i.e. 0916-JAN?

What is your class number?

Teacher-Course Evaluation (TCE) Questions

Instructions: Please respond to the following questions and statements.*

1. What is your overall rating of this instructor’s teaching effectiveness?
   - almost always effective
   - usually effective
   - sometimes effective
   - rarely effective
   - almost never effective

2. How much do you feel you have learned in this course?
   - an exceptional amount
   - more than usual
   - about as much as usual
   - less than usual
   - almost nothing

3. What is your overall rating of this course?
   - one of the best
   - better than average
   - about average
   - worse than average
   - one of the worst

4. Rate the usefulness of the in-class activities (lectures, discussions, etc.) in this course in helping you learn?
   - almost always useful
   - usually useful
   - sometimes useful
   - rarely useful
   - almost never useful

5. I was treated with respect in this class.
   - strongly agree
   - agree
   - uncertain
   - disagree
   - strongly disagree

6. What is your rating of this instructor compared with other instructors you have had?
   - one of the most effective
   - more effective than most
   - about as effective as most
   - less effective than most
   - one of the least effective

By submitting this form you grant permission for the data to be used for research purposes of the current study.

*These questions were taken directly from the official TCE form distributed by the Office of Instructional Assessment, Arizona Board of Regents (2002)
REFERENCES


*Language Learning, 311*, 87-112.


