

**A JAPANESE COLT: ANALYZING TEACHING PERFORMANCE  
IN A JUNIOR HIGH SCHOOL PRACTICUM**

---

A Dissertation  
Submitted  
to the Temple University Graduate Board

---

In Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Education

---

By  
Sadayuki Mitsuo  
May, 2010

Examining Committee Members

David Beglar, Advisory Chair, CITE/TESOL  
Marshall Childs, CITE/TESOL  
Mitsue Allen-Tamai, External Member, Chiba University  
Masako Sakamoto, External Member, Bunka Women's University  
Edward Schaefer, External Member, Ochanomizu University

©

Copyright

2010

by

Sadayuki Mitsuo

## ABSTRACT

The two main purposes of this study were to create a systematic observation instrument in order to obtain clearer and more specific feedback from junior high school teachers about student teachers' teaching performances during their practicum, and to provide a way for junior college, university teachers, student teachers, and practicum supervisors to observe student teachers' teaching and then to communicate their observations more effectively with one another.

The participants were 57 student teachers, 19 college teachers, and 28 junior high school teachers. Four instruments were used: a written consent form, a questionnaire about 15 teaching skills (*The Teaching Skill Questionnaire*), a 60-minute videotape with a checklist (*The Japanese COLT*), and a 42-item questionnaire (*The Student Teachers' Videotaped Instruction*).

The study produced four major findings. First, by using the Japanese COLT (Communicative Orientation of Language Teaching Observation scheme), the three groups of raters (student teachers, college teachers, practicum supervisor) identified four specific problems with individual student teacher's teaching. They (a) explained new sentence patterns without interacting with the students, (b) asked fewer questions than expected, (c) had the students practice reading for a shorter time than expected, and (d) provided few opportunities for the students to speak in Japanese or English, and spoke Japanese more than necessary. The second finding was that the student teachers differed from the older teacher groups in their views of specific teaching skills because of their limited teaching experience and lower English proficiency. The third finding was that the three groups of raters perceived the student teachers' teaching on the videotape similarly. The fourth finding indicated that there was no statistically significant difference in the three groups' views of the teaching techniques used by the student teachers; however, a statistically significant difference was found for the three groups' evaluations of the student teachers' teaching.

The Japanese COLT was a useful instrument for assessing the student teachers' classroom performances, as it provided more specific feedback to the student teachers, and allowed the three groups to share their viewpoints more effectively.

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude towards Dr. David Beglar, my primary advisor who guided me and gave me priceless insight into the writing of my dissertation. I am also grateful to Dr. Mitsue Allen-Tamai, Dr. Masako Sakamoto, Dr. Edward Schaefer, and Dr. Marshall Childs. Each of them gave me helpful comments which aided me in developing my dissertation. I am also grateful to Dr. Steven Ross who gave me the advice concerning how to conduct the research in the initial stage.

I gratefully acknowledge the student teachers, the college teachers, and the junior high school teachers who participated in the study. Especially the comments the junior high school teachers gave me were helpful, as they gave me insight into analyzing the data.

I must also thank Dr. Kenneth Schafer and Dr. Noël Houck, both of whom encouraged me for a number of years to continue working on the study. I also would like to thank Ms. Tomoko Nemoto, the TESOL program Assistant Director, and the office staff of the Graduate Education office at Temple University, Japan Campus who helped me register and take other necessary administrative steps to complete the program.

I would also like to thank my parents: my father, who sent me to England in my youth to be a good teacher of English, and my mother who kept me asking about the progress of my dissertation.

Finally, I must apologize to and thank my wife and daughter, Kyoko and Risako, who were patient enough to let me stay in my study all day long and study on Sundays. I will try to be a better husband and a better father from now on.

## TABLE OF CONTENTS

	PAGE
ABSTRACT.....	iii
ACKNOWLEDGMENTS .....	v
LIST OF TABLES .....	xi
LIST OF FIGURES .....	xiv
CHAPTER	
1. INTRODUCTION.....	1
The Background of the Issue.....	1
Statement of the Problem.....	3
Purpose of the Study .....	6
The Audience of the Study.....	7
Delimitations .....	8
The Organization of the Study.....	10
2. REVIEW OF THE LITERATURE.....	11
Classroom Research.....	11
Experimental Research.....	13
Naturalistic Research.....	16
Observation System.....	32
Conversational Analysis.....	36
Action Research.....	40
A New Approach to Teaching English .....	46

The Need for Process-product Research .....	50
The Validity of the Observation Categories of the COLT .....	52
Inter-rater Reliability in the Use of the COLT .....	54
A New Version of the COLT .....	54
The Use of the COLT in the Classroom Research .....	58
3. TEACHER TRAINING AND OBSERVATION.....	67
Introduction.....	67
A Practicum for Prospective English Teachers in Japan .....	72
Feedback.....	77
Observation Scheme for Providing Feedback to Student Teachers.....	82
Gaps in the Literature.....	83
Purpose of the Study and Research Questions.....	84
4. A NEW MODEL OBSERVATION SHEET .....	86
Phase 1: Revising the Evaluation Sheet.....	87
Phase 2: Developing Additional Criteria .....	92
Phase 3: Student Teachers' Self-evaluation.....	95
Investigating Student Teachers' Practicum in 1996.....	96
Teaching Procedures.....	103
Revision of the COLT for Student Teachers' Practicum at Japanese Junior High Schools .....	112
The Revised COLT .....	116
Other Categories on the Revised COLT.....	119
The Revised COLT and its Categories.....	123

Focus of the Lesson .....	123
Teacher's Activities .....	126
Student Activities .....	128
Rationale for and Further Explanation of the Categories.....	130
Time and Procedures of the Class .....	130
Use of Language in Teacher's Activities, Student Activities, and Other Categories .....	131
The Pilot Study.....	136
Participants .....	137
Materials .....	138
Procedures .....	141
Results .....	142
5. METHODS.....	145
Participants.....	145
Instrumentation .....	166
The Teaching Skill Questionnaire .....	147
The Revised COLT.....	148
The Student Teachers .....	149
The Student Teachers' Videotaped Instruction .....	150
Procedures .....	151
Analysis.....	152
The Rasch Model .....	155
Rasch Rating Scale Model.....	156

6. RESULTS .....	158
Research Question 1.....	158
Research Question 2.....	158
Question 1-5: Teaching Mainly in Japanese .....	159
Question 6-10: Teaching Using Easy English.....	164
Question 11-15: Teaching Using Difficult English.....	170
Research Question 3.....	175
Student Teacher 1 .....	177
Student Teacher 2 .....	179
Student Teacher 3 .....	180
Student Teacher 4 .....	182
Student Teacher 5 .....	184
Student Teacher’s Videotaped Instruction.....	188
Very Useful Or Not Very Useful Worded Likert Items .....	189
The Effectiveness of the Student Teachers’ Instruction .....	204
Differently Worded Likert Scales.....	219
7. DISCUSSION.....	226
Research Question 2.....	226
Construct 1: Teaching Mainly in Japanese.....	227
Construct 2: Teaching Using Easy English .....	233
Construct 3: Teaching Using Difficult English .....	236
Research Question 3.....	241
Research Question 4.....	245
Constructs Related to the Usefulness of the Instruction.....	245

Constructs Related to the Effectiveness of the Instruction .....	251
Additional Likert Items .....	259
Theoretical Consequences of the Results .....	267
Research Question 1.....	272
The COLT and the Revised COLT.....	273
The Role of the Revised COLT.....	278
Problems with the Revised COLT.....	280
Use of English and Japanese.....	281
8. CONCLUSION .....	284
Summary of the Findings.....	284
Limitations .....	290
Suggestions for Future Research.....	293
Final Conclusion .....	297
REFERENCES CITED.....	301
APPENDICES	
A. COLT A AND COLT B .....	323
B. THE STUDENT TEACHERS' PRACTICUM TEACHING QUESTIONNAIRE .....	325
C. STUDENT TEACHERS' ENGLISH PROFICIENCY.....	327
D. GRADE LEVELS AND PERIODS THE STUDENT TEACHERS TAUGHT .....	330
E. SEVENTEEN EXAMPLES OF THE STUDENT TEACHER'S TEACHING .....	331
F. THE RASHOMON EFFECT (JAPANESE VERSION) .....	337
G. THE TEACHING SKILL QUESTIONNAIRE (ENGLISH VERSION) .....	338

H. THE REVISED COLT .....	340
I. THE STUDENT TEACHERS' VIDEO TAPED INSTRUCTION (ENGLISH VERSION) .....	345
J. LETTER OF ACCEPTANCE.....	351

## LIST OF TABLES

Table	Page
1. Revised Evaluation Sheet Criteria.....	88
2. Revised Evaluation Sheet Criteria Descriptive Statistics.....	89
3. A New List of Assessment Criteria.....	93
4. Abstract from the Student Teachers' Practicum Teaching Questionnaire .....	99
5. The Student Teachers' Self-evaluation of their Teaching Performance.....	100
6. While-reading Activities in the 17 Student-Teacher Lessons .....	108
7. The List of Criteria (1996 version).....	111
8. Activities and Episodes on the Revised COLT .....	118
9. The Seven Main Categories and Their Subcategories on the Revised COLT .....	124
10. Comparison of the Answers of the Teaching Skill Questionnaire Among the Three Groups.....	142
11. Revised Rating Scale Category Structure for Teaching Mainly in Japanese.....	160
12. Descriptive Statistics for Teaching Mainly in Japanese.....	161
13. Rasch Difficulty and Fit Statistics for Teaching Mainly in Japanese (Items 1-5) .....	163
14. Revised Rating Scale Category Structure for Teaching Using Easy English.....	165
15. Descriptive Statistics for Teaching Using Easy English.....	166
16. Rasch Difficult and Fit for Teaching Using Easy English (Items 6-10) .....	168
17. Rating Scale Category Structure for Teaching English Using Difficult English.....	170
18. Descriptive Statistics for Teaching Using Difficult English .....	171

19. Rasch Difficulty and Fit Statistics for Teaching Using Difficult English.....	173
20. Descriptive Statistics for the Three Groups' Ratings for Student Teacher 1 .....	177
21. Descriptive Statistics for the Three Groups' Ratings for Student Teacher 2 .....	179
22. Descriptive Statistics for the Three Groups' Ratings for Student Teacher 3. ....	182
23. Descriptive Statistics for the Three Groups' Ratings for Student Teacher 4 .....	184
24. Descriptive Statistics for the Three Groups' Ratings for Student Teacher 5 .....	185
25. Rating Scale Functioning for Listening and Japanese Use.....	190
26. Descriptive Statistics for the Listening and Japanese Use Items.....	190
27. Rasch Descriptive Statistics for the Listening and Japanese Use Items.....	191
28. Rating Scale Functioning for English Use .....	194
29. Descriptive Statistics for English Use. ....	194
30. Rasch Descriptive Statistics for the English Use Items.....	195
31. Rating Scale Functioning for Prereading.....	197
32. Rasch Descriptive Statistics for the Prereading Items.....	198
33. Descriptive Statistics for Prereading .....	198
34. Rating Scale Functioning for Memorization .....	201
35. Descriptive Statistics for Memorization.....	202
36. Rasch Descriptive Statistics for the Memorization Items. ....	202
37. Rating Scale Functioning for Explanation and Understanding .....	206
38. Descriptive Statistics for Explanation and Understanding .....	207
39. Rasch Descriptive Statistics for the Understanding and Explanation Items .....	208
40. Rating Scale Functioning for Practice and Habit Formation.....	211
41. Descriptive Statistics for Practice and Habit Formation .....	211
42. Rasch Descriptive Statistics for the Practice and Habit Formation Items.....	212

43. Rating Scale Functioning for Communicative Practice. ....	215
44. Rasch Descriptive Statistics for the Communicative Practice Items.....	216
45. Descriptive Statistics for Communicative Practice .....	217
46. Descriptive Statistics for Question 18 .....	221
47. Descriptive Statistics for Question 20 .....	222
48. Crosstabulation Results for Question 21 .....	224
49. Crosstabulation Results for Question 22 .....	225
50. Mean Scores of the Three Groups for Items 3 and 4.....	232
51. Mean Scores of the Three Groups for Items 8 and 9.....	235
52. The Mean Ratings (Rasch Logits) for the Three Groups for Students 1-5.....	243
53. Chi-square Table for Item 21.....	264
54. Chi-square Table for Item 22.....	266
55. Mean Scores of the Three Groups for Items 21 and 22.....	267
56. Statistically Significant Differences Between the Three Groups .....	270
57. Teaching Skills in Which the Student Teachers Need Further Training.....	272

## LIST OF FIGURES

Figure	Page
1. Wright map for Teaching Mainly in Japanese .....	162
2. Wright map for Teaching Using Easy English .....	167
3. Wright map for Teaching English Using Difficult English. ....	172
4. All facets vertical rules for student teacher 1.....	178
5. All facets vertical rules for student teacher 2.....	181
6. All facets vertical rules for student teacher 3.....	183
7. All facets vertical rules for student teacher 4.....	186
8. All facets vertical rules for student teacher 5.....	187
9. Wright map for Listening and Japanese Use .....	193
10. Wright map for English Use .....	197
11. Wright map for Prereading. ....	200
12. Wright map for Memorization .....	205
13. Wright map for Explanation and Understanding .....	210
14. Wright map for Practice and Habit Formation.....	213
15. Wright map for Communicative Practice. ....	218
16. The mean Rasch person ability estimates (CHIPS) of each teacher group for the three constructs .....	241
17. The three groups' mean ratings (Rasch logits) for student teachers 1-5.....	242
18. The mean Rasch person ability estimates (CHIPS) of the three usefulness constructs.....	269
19. The mean Rasch person ability estimates (CHIPS) of the three effectiveness constructs.....	269

## **CHAPTER 1**

### **INTRODUCTION**

#### The Background of the Issue

Modern classroom research was developed in the 1950s in the field of general education to help student teachers better understand their approaches to teaching subject matter. Since the 1970s, work has taken place in this same area in the field of second language teaching. The advent of communicative language teaching brought radical changes to many foreign language classrooms. These changes were described and implemented in general terms, such as the provision of comprehensible input and a focus on meaning; however, there were also many interpretations of communicative language teaching and considerable variance in the ways in which teachers implemented it. Under these circumstances, it was necessary for researchers to observe foreign language classrooms in order to identify differences in the implementation and effectiveness of these different approaches. For this reason, a variety of observation schemes were developed. One of the best known of these schemes, the Communicative Orientation of Language Teaching Observation scheme (COLT) (Allen, Fröhlich, & Spada, 1984), was

designed to help researchers describe classroom processes and examine these processes in relation to learning outcomes.

In Japan, the Japanese Ministry of Education (1989, 1990) introduced new foreign language courses into public junior and senior high schools nearly two decades ago, and as a part of these changes, high school English teachers were required to adapt their teaching to the communicative language teaching approach. Not surprisingly, these changes impacted the training and evaluation of student teachers who participate in a three-week teaching practicum in junior or senior high schools as part of the teacher licensure program when they are university seniors. As one of the main requirements of the program, the student teachers are required to teach English. Their teaching performance is evaluated and the results are sent to their college advisors after their completion of the practicum. Evaluation sheets detailing the students' performance are then prepared by their junior college or university advisors. These college evaluation sheets, which are made up of items that reflect the standards set by the Ministry of Education, Science and Culture, include: (a) teaching instruction, which is made up of three subcategories, teaching plans, teaching techniques, and evaluation of teaching; (b) management of the class and students, and; (c) overall competence, aptitude, and working attitude.

Practicum supervisors also write comments and award each student teacher a holistic grade ranging from 1 (the lowest grade) to 5 (the highest grade) (Anonymous, 1993; Kurosaki, 2002; Mitsuo, 2000). The grading criteria are not standardized; thus, a large number of criteria are used by various supervisors (Anonymous 1993; Kurosaki 2002; Mitsuo, 2000).

### Statement of the Problem

The first problem addressed in this study concerns the lack of clarity in the feedback provided to student teachers. They must obtain clear feedback about their performance during their practicum if they are to learn more about their strengths and weaknesses as a teacher, and reflect on their teaching with greater awareness. One opportunity to do so occurs when they observe and critique their teaching on videotape. Clearer criteria and higher quality feedback would permit all parties involved—the student teachers, university teachers, and junior high school teachers—to have more constructive discussions.

The second problem is that the junior and senior high school supervisors' feedback regarding the student teachers' teaching performance is generally vague, short, and unsystematic. As a result, the junior college teachers and university

teachers involved in teacher licensure programs can not adequately understand the overall quality of the student teachers' performance or the shortcomings in their performances. In addition, their comments are not particularly specific; thus, junior college or university teachers have difficulty responding adequately to the supervisors and making good use of the feedback when they instruct the student teachers in their English teaching methodology courses.

The third problem is that grading varies considerably across different educational contexts in Japan. Different supervisors and educational institutions emphasize different criteria when awarding grades to student teachers' practicum performances; this practice can result in a lack of reliability of the grades that the student teachers are awarded as well as a lack of comparability.

The fourth problem is that grading varies from lenient to strict depending on the supervisor. This problem is extremely serious as the practicum grade strongly affects the part-time teachers' employment possibilities in the future. The teaching performance evaluation in the practicum is part of the hiring criteria when new graduates are employed as part-time teachers; thus, any unfairness in these evaluations can impact a student's chances for employment. Clearer and more

consistent criteria will potentially result in more systematic assessment that is fairer to the student teachers.

The fifth problem concerns the current system for employing primary and secondary school teachers in Japan. For example, in Tokyo, some university students studying in a college of education that provides a teacher licensure program for elementary school teachers can teach as an assistant teacher in their fourth year for one year and obtain a full-time teaching position the following academic year if their performance as an assistant teacher is evaluated highly. This system has been implemented for two reasons: to employ talented students and to employ student teachers after evaluating their teaching performance in a classroom context rather than from paper-based tests and interviews. However, this employment process is controversial in Japan. If the teaching performance evaluation system is made more systematic and its criteria clarified, it will be possible to choose new teachers based on their teaching performance in a longer practicum rather than through the use of paper-based tests and short interviews.

## Purposes of the Study

The first purpose of this study is to create a systematic observation instrument in order to obtain clearer and more specific feedback from junior high school teachers about student teachers' teaching performances during their practicum. Systematic criteria will help supervisors provide clearer and more specific feedback; these criteria can also be shared by supervisors across junior high schools as well as by practicum supervisors and college teachers. This will allow administrators, student teachers, junior and senior high school teachers, and the student teachers' college advisors to share the same language when they communicate about the student teachers' teaching performance.

The second purpose of this study is to provide a way for junior college or university teachers, student teachers, and supervisors to observe student teachers' teaching and then to communicate their observations more effectively with one another. People in different positions and with different degrees of education and experience frequently have different points of view. While a variety of viewpoints is important for improving the educational process, it is also necessary for them to share the same point of view as it constitutes a common ground from which they can discuss the student teachers' teaching constructively.

## The Audience for the Study

The first audience for this study is university professors who are involved in teacher licensure programs. The results of this study potentially provide them with ideas for developing more effective evaluation criteria, evaluation instruments, and approaches to data analysis. University professors can obtain specific ideas concerning what student teachers should have studied before their practicum, so they can better consider what to teach in their teaching methodology courses.

The second audience is student teachers who have returned from their practicum. The use of a well-organized, clearly-written observation sheet will permit student teachers to evaluate their own teaching more effectively as they recall their teaching experiences and watch videotapes of their classroom performance. Affording student teachers a clearer understanding of their strengths and weaknesses also allows them to discuss their performance with their university advisors and junior or senior high school supervisors more effectively.

The third audience is the practicum supervisors at the junior high schools. An observation sheet with detailed criteria based on the principles underlying Communicative Language Teaching will permit junior and senior high school teachers to provide better feedback concerning the student teachers' teaching

performance, while also familiarizing them further with communicative language teaching principles

The fourth audience is the members of local boards of education. If junior high school supervisors are able to assess student teachers' performances using criteria that are more specific and clearer than those currently in use, the grades might be more reliable. Moreover, if the criteria are shared by supervisors across many schools, the grades will be more interpretable and potential teachers more easily compared. This will benefit local boards of education because they select new part time teachers partly based on the grades that those student teachers receive in their practicum.

#### Delimitations

The problems addressed in this study exist in many teacher training programs currently operating in Japan; thus, the findings of this study can be generalized to other teacher training programs in Japanese universities. The primary delimitation concerns the applicability of the results to teacher training programs in other countries. While there may be similarities between the educational context described in this study and those in some other countries, the

results can only be generalized to non-Japanese contexts with considerable caution; all participants in this study are Japanese and the educational system and approach to teacher training described in this study likely differ in important ways compared to those in other countries; thus, it is inadvisable to draw strong conclusions from this study regarding the assessment of student teachers taking part in teaching practicum in non-Japanese contexts.

The second delimitation concerns the participants' English proficiency level. Generally speaking, the student teacher participants in this study had low English proficiency, so the degree to which the findings of this study can be generalized to students with higher English proficiency is unclear. Also, because all of the student teacher participants came from only two colleges, it is not clear how well the results can be generalized to students studying in other teacher licensure programs. In addition, all of the participants in this study had experience in practice teaching to a group of their peers at the university. However, this practice is not universal and many student teachers from large universities and colleges do not have any practice teaching experience. Thus, the results of the study might not generalize well to those students. This delimitation does not affect the discussion of ways

junior high school teachers can provide more specific feedback to student teachers and college teachers involved in college teacher licensure programs.

### The Organization of this Study

Chapter 2 is a review of the classroom observation literature, including observational methods such as observation schedules, questionnaires, and ethnographic methods. Chapter 3 is literature review of teacher training and observation. The literature review is followed by a description of the gaps in the literature and the research questions that guide this study. A new model observation sheet is presented and discussed in Chapter 4. A description of the participants, the instrumentation, and procedures used in this study is provided in Chapter 5, the Methods Chapter. The results for each research question are provided in Chapter 6, and those results are discussed in Chapter 7. In Chapter 8, Conclusion, I present the limitations of the study, make suggestions for future research, and provide brief concluding comments.

## **CHAPTER 2**

### **REVIEW OF THE LITERATURE**

#### Classroom Research

Modern classroom research was developed in the 1950s in order to help student teachers better understand their approaches to teaching subject-matter classes. The data gained through classroom research provided them with feedback about their classroom performance. From the viewpoint of teacher trainers, it was important to know the defining characteristics of effective teaching and incorporate empirical findings into teacher training regimes in order to make them more efficient and effective. One example of work in this area was carried out by Flanders (1970), who developed an observation system based on a list of teacher and learner behavior categories thought to be the most useful in promoting effective teaching and learning. These observation sheets were used to help teacher trainees understand their teaching methods more clearly.

Classroom research has been conducted with a wide range of research methods selected because of their usefulness in illuminating classroom language learning and teaching. In classroom observation, which is one of the major methods

used by researchers conducting classroom research, researchers observe what occurs in classrooms in order to relate the major features of teacher and student behavior to learning outcomes.

Researchers can develop a database of classroom activities by direct observation, note-taking, and video- and/or audio-recording classroom behavior and interactions. Verbal interactions that are recorded can be transcribed and notes concerning nonverbal behaviors can be added to the transcripts. However, other means of classroom research must be used in order to investigate aspects of the classroom that cannot be observed directly. For example, if the aim is to investigate learners' anxiety, one alternative to direct observation is to interview the students about what has occurred in the class and how they feel about it, or to administer written questionnaires. A problem with observation systems and questionnaires is that the researcher must prejudge the classroom in order to select criteria for inclusion on the observation instruments. While observation instruments and questionnaires can yield valuable data, an alternative approach to data collection is to use an open form of self-report. For example, learners' diaries can reveal aspects of the classroom experience that direct observation might miss and that researchers would not have thought of including as questionnaire items. These limitations of

observation schedules and questionnaires led some researchers to use ethnographic methods to collect and interpret classroom data. The use of ethnographic methods has received support from many researchers in general education for over three decades (Erickson, 1977; van Lier, 1988; Wilson, 1977). The value of ethnographic approaches was pointed out by Van Lier, who stated that researchers must attempt to understand the meanings given to the classroom events by the participants themselves.

Classroom research techniques have been developed systematically in the field of second language (L2) teaching, and researchers studying L2 classrooms have been influenced by a number of disciplines including education, sociology, psychology, linguistics, and applied linguistics. Three major research approaches are evident in the development of classroom research in the classroom: experimental research, naturalistic inquiry, and action research. Those three approaches are reviewed in the following sections.

### *Experimental Research*

Experimental research was applied in early evaluations of L2 instruction in the 1950s and 1960s. A number of researchers followed standard educational

psychometric procedures closely with comparison treatment groups and the measurement of outcomes on proficiency tests. Examples of these standards include Scherer and Wertheimer (1964) and the Pennsylvania project (Smith, 1970). These major studies, which were conducted in order to determine the efficacy of the main teaching methods of that time, proved inconclusive. For instance, Scherer and Wertheimer compared the audiolingual method with the traditional grammar-translation teaching at the university level and found no significant differences overall. In the Pennsylvania project (Smith, 1970), audiolingualism was compared with traditional teaching, which was defined as grammar-translation. This experiment also failed to show any statistically significant differences between the results of several tests.

In later studies, the focus was transferred from teaching methods to teaching techniques, but the research paradigm remained the same: Researchers attempted to understand effective teaching by comparing teaching techniques. For instance, according to Lindblad (1969), researchers involved in the Gothenburg English Training Method Project (GUME) tested the usefulness of grammatical explanations that were based on Chomsky's 1957 version of Transformational-Generative Grammar. With child participants, the comparison was inconclusive:

Children learning from explanation and practice did no better than those learning only by practice. In the case of adult participants, however, grammar explanation and practice resulted in greater learning than practice alone, but the researchers could not generalize the findings beyond their sample because of the small scale of the GUME project in terms of the number of lessons involved, the number of teaching points covered, and the fact that the teaching was on audiotape.

Politzer (1970) also compared certain instructional techniques in secondary school French classes. These pedagogical techniques mostly involved different types of structural pattern practice. He recorded the frequencies of those techniques and related the frequencies to learner achievement in different classes. The results were complex and interesting, yet Politzer admitted that “the very high complexity of the teaching process makes it very difficult to talk in absolute terms about “good” and “bad” devices (p. 43).

Gritter (1968, p. 7) concluded that “...perhaps we should ask for a cease-fire while we search for a more productive means of investigation.” This conclusion is shared by more recent researchers, such as Larsen-Freeman (1996, p. 63), who pointed out, “researchers have come to recognize the limits of process-product research in helping us to develop an understanding of teaching and learning.”

Gritter's comment was partly responsible for the move towards the second classroom research approach, naturalistic inquiry, though some experimental research was conducted after 1970. For example, Bejarano (1987) used an experimental approach to investigate cooperative group work in language classrooms in Israel. He reported on the effects of two small-group cooperative techniques (Discussion Group; Student Teams, and Achievement Divisions) and the whole-class method on EFL academic achievement for 665 pupils in 33 seventh grade classes. The findings revealed that both group methods resulted in significantly greater improvement than the whole-class method on the total test score and on the listening comprehension scale. These findings support the link between the communicative approach to foreign language instruction and cooperative learning in small groups.

### *Naturalistic Inquiry*

Research based on naturalistic inquiry brought about two changes in classroom research. The first was that researchers' focus turned from a prescriptive to a descriptive approach that was used with naturally occurring settings and groups. The second change was that the focus moved from teaching methods and

techniques to the process of teaching and learning. These changes meant that the researchers had to find ways of describing classroom processes: how the class proceeds, and what teachers and students do in the classroom. Naturalistic inquiry includes a number of different methodologies, including ethnographies, case studies (including diary studies), and more general observational studies. Researchers have used a variety of coding systems as well as conversational analysis to analyze the data collected through these methods.

### *Ethnographic Research.*

Ethnography is “concerned primarily with the description and analysis of culture” (Saville-Troike, 1982, p. 1) and is “the study of people’s behavior in naturally occurring, ongoing settings, with a focus on the cultural interpretation of behavior” (Watson-Gegeo, 1988, p. 576). The ethnographic tradition is generally identified as a qualitative, process-oriented approach to the study of interaction. It has been developed in many ways by L1 classroom researchers (Barnes, Britton, & Rosen, 1969; Cazden, 1986; Cazden, John, & Hymes, 1972; Chaudron, 1980; Wilkinson, 1982) and has been employed by L2 researchers to a limited degree in part because it requires highly trained skills and a great deal of time and

commitment by the researchers. Continuous record keeping, extensive participatory involvement of the researcher in the classroom, and careful interpretation of the data gathered in the class are required. Such an investigation usually leads to a precise description of the site as well as the rules used among the participants as they interact with each other.

As a research method, ethnography is most often associated with anthropology. However, it has also been productively utilized in studies of language education. One early example of ethnographic research conducted in a foreign language classroom is Cleghorn and Genesee's (1984) report on a French immersion classroom in Canada. More recent examples of classroom ethnographies are van Lier's (1996) study of a bilingual program in Peru, Duff's (1996) work with dual-language, late-immersion secondary school programs in Hungary, and Lin's (1999) comparison of four English classrooms in Hong Kong.

van Lier (1996) described the language use of children and teachers in a Spanish-Quechua bilingual education program in Peru. He presented a vivid picture of an attempt at educational innovation, along with his concerns about whether the program and its accompanying research agenda could be sustained over time. In the rural communities of the Altiplano, where van Lier worked as a teacher and

researcher, most of the children are monolingual in Quechua or Aymara when they enter the first grade. All schooling has traditionally been conducted in Spanish, with varying degrees of tolerance of the native language in the first three grades. During the reign of the revolutionary government in the late 1960s and early 1970s, there was a strong push for the revitalization of indigenous languages and cultures, and in 1975 Quechua was declared an official language alongside Spanish. The PEEB project (Proyecto Experimental de Educación Bilingüe), in which the goal was to maintain the children's native language throughout elementary school, was proposed and implemented around 1980. van Lier was involved in the overall monitoring and evaluation of the project for the two years he was there. He frequently visited communities in which the program was implemented in one or more elementary classrooms. In particular, he often spent an entire week at two schools assessing the effectiveness of the project on a longitudinal basis. One of these schools, Tiyaña, was a project school in which bilingual education had been implemented. The other, Qotokancha, was a "comparison school" in which there were no bilingual grades. During these visits he administered entry and exit tests to all the children in the bilingual and comparison schools, both in Quechua and Spanish (spoken and written), observed classes, and talked to teachers, parents, and

students about many issues, pedagogical and otherwise. He also played volleyball and fulbito (a kind of soccer) and attended community meetings. Through these contacts with people in the community and observations of the classrooms, he described the project implemented in these areas.

Duff (1996) investigated the socialization of discourse competence in two instructional environments in Hungarian secondary schools. The first was a traditional monolingual school in which the traditional pedagogical strategy called 'felelés' (a recitation) is dominant. The second was a dual-language school in which the instruction took place mainly in English. Duff's broad goal was to analyze the impact of the massive social changes wrought within the educational system with the end of Soviet domination in Hungary. The data included approximately fifty videotaped lessons, as well as written and oral comments from teachers and students. Duff used her data to highlight issues of educational and linguistic reform in a rapidly changing political environment.

The felelés constituted the standard means of assessing students' progress in their content classes. Originally, its purpose was to develop students' moral character; secondary goals that are still upheld today are to foster discipline, patriotism, conformity, oral self-expression, and the accumulation and review of

knowledge presented in class. After a felelés, no students are expected to comment or ask questions. Teachers award a grade on a 1 (the lowest score) to 5 (the highest score) scale.

Students at a dual language school perceived large differences between the two approaches—the freedom and democracy they experienced at the DL schools, and most dramatically in the entry-level year, compared with primary schools where the felelés is dominant. As a consequence of their successful EFL learning experience in the free atmosphere of the DL schools, the students became more interactive and more likely to express themselves in class as well as more demanding about school practices and opportunities to further their academic goals and other aspirations. Because of the success of the approach used in the DL school, the felelés has lost its luster. Under the domain of the felelés, students had to understand what they would say in front of the class and memorize it. At the DL schools, a ‘lecture’ was given by a student instead of a felelés. During the lecture, the students were able to look at notes. For this reason, and probably because it reduced preparation time in comparison to the time that they had to spend for a felelés, the students often read rather than reciting or informally discussing issues when it was their turn to lecture.

Lin (1999) described teachers' discourse structures in four classrooms situated in different socioeconomic settings in her attempt to focus on the classroom dilemmas in which the students and teachers found themselves. For example, Teacher A's discourse structure was *teacher initiation* (L2-L1) followed by *student response* (L1) followed by *teacher feedback* (L1-L2). In this class, the students were not required to reformulate their L1 responses in the L2, as the teacher did it for them in the feedback slot of the IRF format. In the case of Teacher B, three structures were identified. One structure was adopted for story-focus: *teacher initiation* (L1), followed by *student response* (L1), followed by *teacher feedback* (L1). Another structure was focused on language: *teacher initiation* (L1 or L2) followed by *student response* (L1 or L2), followed by *teacher feedback* (L2, or restart with teacher initiation (L1 or L2) until student response is in L2). The other structure was to start along the previous discourse structure again to focus another linguistic aspect of the elicited L2 response, or to return to the first structure to focus on the story again.

Lin drew on three notions: (a) *cultural capital*, the language use, skills, and orientations, dispositions, attitudes, and schemes of perception that children are endowed with by virtue of socialization in their families and communities

(Bourdieu, 1984, 1991); (b) *symbolic violence*, which concerns how the disadvantaging effect of the schooling system is masked or legitimized in people's consciousness (Bourdieu, 1984), and (c) *creative, discursive agency*, which is the strategies that people use to cope with these dilemmas. The notion of creative, discursive agency (Collins, 1993) is rooted in the phenomenological tradition that stresses the creative, emergent practices of social actors, who are not simply puppets of larger social forces and structures. Lin discussed the possibility of *creative, discursive agency* by referring to Teacher B's teaching. In Teacher B's classroom, the students came from a disadvantage socioeconomic background and their habitus did not equip them with the right kind of attitudes, interest, skills, or confidence in learning English. However, there were signs of their habitus being transformed through the creative, discursive agency and efforts of Teacher B. For example, she used the L1 strategically in the reading lesson to intertwine an interesting story focus and a language learning focus. She helped her students experience a sense of achievement and confidence in learning English. At school, she spent most of her spare time with her students establishing a personal relationship with each of them. With all these extra personal, creative efforts, she succeeded in helping her students develop greater interest, skills, and confidence in

learning English. Lin implied that understanding existing classroom practices and their sociocultural and institutional situatedness is a first step towards exploring the possibility of alternative creative, discursive practices that might contribute to the transformation of the students' habitus.

### *Case Studies*

Another type of naturalistic inquiry in second language acquisition research is the case study. When conducting a case study, "one selects an instance from the class of objects and phenomena one is investigating (for example, 'a second language learner' or 'a science classroom') and investigates the way this instance functions in context" (Nunan, 1992, p. 75). One well-known example of a case study in the field of second language acquisition is that of Schmidt (1983), who conducted a longitudinal case study of Wes, an adult learner of English in Hawaii. He described how Wes improved his English and why part of his English fossilized. This case study was conducted outside of a formal classroom environment, but the case study approach can also be used in formal instructional settings as well. A good example of this approach is classroom research by Donato and Adair-Houck (1992). They reported on two secondary school teachers' lessons

of the French future tense. The two teachers, Elizabeth and Claire, displayed markedly stable but different strategies for teaching the future tense, a process that took eight lessons for Elizabeth and ten for Claire. After videotaping, transcribing, and analyzing the lessons, the researchers described the teachers' approaches using excerpts from the transcription: Elizabeth's orientation being monologic and Claire's being dialogic. Elizabeth chose topics and spoke to the students, so the students had few opportunities to speak. In contrast, Claire encouraged the students to respond to her when she initiated a topic, she was responsive to the students' contributions, and she was comfortable letting the students initiate talk.

Cotterall (2004) conducted a case study with Harry, a 29-year-old native speaker of English enrolled in his first year of study towards a Bachelor of Arts degree at Victoria University of Wellington after spending several years as a chef. The goal of the study was to explore the learner's goals and beliefs about language learning as part of his ongoing experience studying Spanish during a 12-week course. Cotterall had six interviews with him over a four-month period in which she asked open-ended questions at the beginning of each session. She found that Harry's focus was narrowly focused on the memorization of grammatical rules throughout the course. At first his interest in the language was motivated by a

desire to learn about the culture, history, and ideas of the Hispanic world, and his specific goals were to acquire the ability to use the language to express himself and to explore the culture of the people who spoke the language. The interviews provide evidence of a consistent narrowing of Harry's goals until the agenda of the course dominated, forcing him to reduce his focus. She concluded that Harry's language learning experience highlighted the necessity of personal importance and that learners' contributions to the curriculum—in terms of goals, interest, and effort—must be not only acknowledged but also utilized in order for the classroom experience to be meaningful.

When using a case study approach, ethnographic researchers usually have provided analyses of specific areas of interaction rather than a complete ethnography of the classrooms that they have observed. Some examples are teacher awareness of student performance (Carrasco, 1981), turn-taking and repair (van Lier, 1982), and teacher management of turns (Enright, 1984). Although these studies did not provide an exhaustive treatment of the rules for interaction in general, the researchers were able to reveal some of the underlying social norms for interpreting specific interactive events in the classrooms they observed.

Wong-Fillmore (1980) conducted a large-scale study of bilingual instruction involving longitudinal participant observation in order to investigate differences between classes in which second language learning went well and did not go well in addition to observing how the teachers influenced the children. Wong-Fillmore observed four classes for one year and saw significant differences among the students in the four classes in terms of their English proficiency independent of ethnicity and native language background. She videotaped the classes, focusing on 19 children who spoke Chinese or Spanish as their first language. She found common characteristics among the successful language classes and the successful teachers. For example, the class activities in the successful classes were consistent and clear for the children and the successful teachers focused on communication and the children's understanding. The study led to several reports of specific desirable qualitative aspects of second language classrooms including functions of language use (Cathcart, 1986) and teacher structuring of input (Wong-Fillmore, 1985). The study also led to quantitative analyses of frequency of interactions, language use, and achievement outcomes.

### *Diary Studies*

Another type of naturalistic inquiry concerns a type of ethnography known as diary studies. These studies (e.g., Bailey & Ochsner, 1983; Brown, 1985; Campbell, 1996; Leung, C-Y, 2002; Schumann, 1980) often involve the researcher-as-learner: (a) recording events in a language classroom; (b) reflecting on diary entries and adding appropriate interpretations soon afterward, and; (c) compiling and summarizing key elements of the diaries and interpretations. Although this approach is relatively subjective, this type of ‘direct’ analysis can provide valid insights if the interpretation of the diaries is based on independent theory and research or the diaries are interpreted with input from other experts and participants (‘indirect’ analysis).

In the past decade, a number of diary studies using indirect analysis have been published (e.g., Allison, 1998; Malcolm, 2004; Sataporn & Lamb, 2004; Umino, 2004). Some studies, such as the one by Allison, were focused on matters of language and course content, while others, such as the study by Sataporn and Lamb, were focused on affective issues and learners’ perceptions of their own language learning behavior, and yet others, such as Malcom’s, were focused on learning strategies.

Allison (1998) focused on the use of language course diaries by looking at diaries kept by 38 second-year undergraduates during an English language course at the National University of Singapore. The author presented an investigation into the use of course diaries as a means of language exploration that can enhance learners' language awareness. The study was focused primarily on matters of language and language content; a preliminary content analysis of the course diaries, learners' responses to a questionnaire, an illustrative account of learners' engagement with language issues in their dairies, and a commentary on teacher feedback and learner reaction were presented. In the preliminary content analysis, the author described the participants' dairy entries. The open-ended spoken guidelines given in lectures to the students about keeping diaries had noted such possibilities as analyzing texts of the students' own choosing, or commenting and raising questions about course readings and tutorial activities. The participants' responses to the questionnaire indicated that many students had completed the work to satisfy the course requirements rather than for intrinsic reasons, and they also acknowledged that they had not kept diaries regularly. In the learners' engagement with language issues, the researcher illustrated some of the ways in which the learners engaged with concepts and analytical procedures that were introduced in

the open-ended spoken guidelines. Sixty-six explicit questions on 22 topics were asked over the year. The topics most frequently raised were prepositions, case grammar categories, homonymy, and polysemy. In the section of the commentary on teacher feedback and learner reactions to the feedback, feedback was provided on issues such as overviews of teaching points that the students had asked about, making references where possible to the students' own examples for discussion. There were precise answers to specific questions, for example, "No, 'asymmetric' does not correspond to 'intransitive.'" There was also an emphasis on the value of asking questions and accepting that some of the answers might not be clear. In the conclusion, he discussed the limitations of the study, emphasizing that language teaching researchers should seek to establish generalizability to other contexts.

Malcom (2004) stated that how learners' beliefs evolve into personal theories of effective language learning is not well documented, although researchers have stressed that learners' beliefs are inherently unstable. He conducted a longitudinal case study with an Arabic student named Hamad in which he detailed his progress over several years and described how his strongly held belief in the value of reading as the key to language development came about. Malcom stated that the study was not static but was modified and refined in relation to changing contexts

and experiences. He also discussed the learner's beliefs and practices in relation to other case studies.

Sataporn and Lamb (2004) described the learning behavior of students taking a self-instructional distance English program at a university and attempted to identify factors that affected their behavior, including their continued participation. The informants, who were attending a one-year Certificate in English for a Specific Career Program, were asked to write 'study diaries' in order to record the regularity and thoroughness of their study habits. In addition, semi-structured interviews were conducted with each participant at the beginning and end of the six-month period. The informants were mostly pleased to have the opportunity to discuss their work and progress. Although the notes in the study diaries tended to be rather superficial, they did provide a means of cross-checking the information from the interviews.

Umino (2004) explored the experience of 20 Japanese learners studying a second language through self-instruction using broadcast materials. She attempted to illuminate the manner in which the learners pursued self-instruction at home using their diaries and interviews. She relied on in-depth interviews with the participants rather than their dairies as the participants did not keep their dairies very well. She identified three factors that contributed to persistence in learning

with the broadcast materials. First, she pointed out the importance of routine setting: Learners who listened to or watched the series at a fixed time, pace, and place were more successful than those who did not do so. Second, learners who started at a younger age received support from their families in one form or another, so they continued to study. These learners were also likely to set long-term goals. Third, the relationship between effort and persistence was an important factor for learners to continue to study.

### Observation System

Most researchers who adopt qualitative or ethnographic techniques have recognized that they also need to adopt quantitative methods. For instance, phenomena that have been counted or measured include the frequency of turns or other units of participation (Allwright, 1980), the frequency with which certain language functions are produced (Cathcart, Strong, & Wong-Fillmore, 1979), and the duration of activities (Mohatt & Erickson, 1981).

Observation systems were originally used to classify teachers' behavior in teacher training, so the focus was on teachers rather than on learners. Although these systems were originally devised for researchers and teachers to observe

classes, the focus shifted from teachers' behavior to both teachers and students' behaviors. These two changes led to the further modification of these observation systems appropriate to the complexities of teaching and learning.

Flanders' 1970 pioneering work in interaction analysis was designed for general education purposes. The main idea underlying interaction analysis was that teaching was more or less effective depending on how 'directly' or 'indirectly' teachers influenced learner behavior. Based on this idea, Flanders produced ten categories that allowed researchers to observe and record both direct influences (e.g., 'criticizing or justifying authority') and indirect influences (e.g., 'accepting learners' ideas'). Teachers, whose teaching was observed, were given scores reflecting the 'directness' or 'indirectness' of their teaching styles.

Moskowitz (1971) modified the categories of Flanders' Interaction Analysis observation system and called this modified version Foreign Language Interaction (Flint). The Flint was used both as a research tool and as a feedback tool in teacher training. Observers using the Flint filled in a matrix specifying several analytical categories. For instance, the category *Teacher talk* was made up of three subcategories, *direct talk*, *indirect talk*, and *student talk*. *Direct talk* included subcategories such as *deals with feelings*. *Indirect talk* included the subcategories

of *gives information* and *gives directions*, while the subcategories of *student talk* included *student response, specific* and *student response, choral*. Entries were made in the matrix during class at regular intervals so that by the end of the lesson a graphic record of events was available. The advantages of this observation system were that no audio or video recordings were made and a large amount of time did not need to be spent transcribing the data. With this tool, student teachers could analyze their own teaching in order to gain objective feedback and a firmer basis for comparisons in their later attempts to teach differently.

In addition to the systems described above, Fanselow (1977) modified and elaborated an analytical system produced by Bellack, Kliebard, Hyman, and Smith (1966) and produced the Foci for Observing Communications Used in Settings (FOCUS), an observation schedule for language teacher training. The FOCUS was made up of five categories: *Who speaks, pedagogic purpose, medium used, area of content, and how mediums are used to communicate content areas*. No separate categories were created for teachers and learners; thus, the categories can be used regardless of the participants and their role in the interaction.

Instruments created for teacher training purposes are not necessarily appropriate tools for some types of classroom research. For example, four

researchers working in Mexico (Long, Adams, McLean, & Castaños, 1976) wanted to investigate the language produced by university level Spanish-speaking students of English under two conditions: in full classroom interaction and in dyads. They found that no instruments developed in the second tradition were appropriate for their research. They needed a system that provided a focus on the communicative variety of speech systems produced by their learners, so they created a new classification system called the Embryonic Category System. This system was used to code the communicative variety of speech systems produced by their learners into three categories: *pedagogical moves*, *social skills*, and *rhetorical acts*.

*Pedagogical moves* was made up of ten subcategories, including *Student initiates discussion*, *Student focuses discussion*, and *Student clarifies*. *Social skills* was comprised of 13 subcategories, including *Students competes for the floor*, *Students interrupts*, and *Students confirms*. *Rhetorical acts* included 14 subcategories, such as *Student predicts*, *Student hypothesizes*, and *Student makes an observation*.

Other problems have been identified with the observation systems discussed above. One is that the categories included in these systems are not the same, so researchers can not use more than one system at a time and compare the observation results (Chaudron, 1988). Another problem is that the categories that

form the unit of analysis in these systems is not defined sufficiently clearly, so researchers can interpret the categories differently. Many researchers are concerned over the potential invalidity of the category systems because each researcher or team chooses to adopt slightly different dimensions and categories, depending on the purposes or theoretical orientation of the study; this often leads to results that are difficult, if not impossible, to compare across studies. These category systems also require that researchers observe classroom interactions using prejudged criteria.

### *Conversation Analysis*

Because of the problems mentioned in the previous section, some researchers turned to transcriptions of recorded classroom events as their primary data source. While producing transcriptions is a time-consuming process, it provides a detailed account of the linguistic interactions that occur in classrooms and the data can be subjected to conversation analyses (Richards & Schmidt, 1983). This procedure includes the detailed microanalysis of such conversational features as socialization, repair, in-breaths, vocalized filters, hesitations, and turn-taking. The analysis approach helped researchers to develop an awareness of the internal formal

structure and functional purposes of verbal classroom interaction. Specific types of discourse phenomena in the classroom (e.g., turn-taking and repair) have required the use of other research methods from the ethnographic tradition. The L1 classroom research of Bellack et al. (1966) is the primary early example of this tradition in education.

Sinclair and Coulthard (1975) built on this approach by developing a system of units that were intended to characterize the functions of pieces of discourse. Sinclair and Coulthard's analysis of transcripts of British elementary classroom verbal interaction allowed them to draw up a hierarchy of units of interaction. They used both linguistic and sociolinguistic traditions in their conception of classroom interaction as a hierarchically structured system of ranks. Their largest unit was the lesson itself. The lesson was made up of *transactions*, each of which consisted of *exchanges*, each of which was made up of *moves*, which consisted of the smallest interactional units, *act*. Acts could be further analyzed into linguistic units like word and phrases.

Second language classroom researchers have not employed a comprehensive discourse analytical scheme in their studies; instead, they have limited themselves

to specific areas of discourse, such as the analysis of teacher feedback (Chaudron, 1977; Tsui, 1985) and adult ESL classroom interaction (Ulichny, 1996).

Ulichny (1996) investigated interaction in an intermediate adult ESL conversation class over a period of two months. The researcher attended classes weekly for six weeks, tape recorded most of the sessions, and transcribed several instances of patterns that she had identified. She found the same teacher-dominated feature in transcriptions gathered across the various patterns. The teacher did most of the talking and determined the size, shape, and nature of one student's contribution to her own story. The most extended conversational sequences included teacher questions requiring a simple yes or no or one-word answer from the student. Two types of discourse activities were dominated by the teacher: In the first type, the teacher corrects the student's English and this correction repairs the conversation at hand and is directed primarily at the participant; In the second type, the instructional activity provides a metadiscussion about the correction and addresses the whole class as language learners. The teacher puts the original conversation on hold for either a correction-by-repetition routine or an instructional routine, which is in turn embedded in the repetition routine. She argued that this feature is clearly teacher constructed and hence unique to this particular teacher and

yet she stated that the basic feature of interrupting an ongoing activity to focus on the language form that students produce is commonplace in ESL classrooms, referring to the positive and negative role of corrective feedback in language instruction. The microanalysis of the interaction showed how one teacher managed the dual pressures of providing authentic language experiences plus structured grammar and vocabulary practice within a single classroom speech event. She concluded with an evaluation of the effectiveness of this type of interaction for language learners and recommended engaging teachers in microanalyses of classroom interaction in order to improve pedagogical practices in L2 classrooms.

Naturalistic inquiry provides classroom researchers with several advantages. First, it permits an in-depth study of individuals, settings, and interactions. As it includes both emic and etic perspectives, it promotes a consideration of all points of view. Second, naturalistic inquiry can address many language issues that are often lost in statistical analyses associated with experimental studies. For example, if the aim is to investigate learners' anxiety, one alternative to direct observation is to simply interview the students about what has occurred in the class and how they feel about it, to administer written questionnaires, or to study learners' self-reports in the form of diaries.

Naturalistic inquiry also has disadvantages. Data collection, data reduction, and data analyses are extremely labor-intensive and time-consuming, particularly because ethnographies, diary studies, and case studies are usually longitudinal. Another disadvantage is the absence of agreed-upon criteria for determining the significance of the outcomes. In the naturalistic approach, generalizability is not always a prime goal; van Lier (1988) argued that generalizability cannot be a major goal because “the first concern must be to analyse the data as they are rather than to compare them to other data to see how similar they are” (p. 2). Thus, the goal in the naturalistic approach is to understand what occurs in the individual classroom, which is a potentially unique social context. Any particular classroom may be more or less similar to other classrooms, but understanding the interaction must precede generalizing its patterns to other settings. In other words, the validation of agreed-upon criteria for determining the significance of outcomes is necessary if researchers are to generalize their findings to other contexts, but this is notoriously difficult to achieve.

### *Action Research*

The third major approach to language classroom research is action research.

While experimental research is often directed at hypothesis testing and theory building, and naturalistic inquiry aims to describe the phenomena under investigation, action research has a more immediate, practical focus. The term *action research* refers to a reiterated cycle of procedures. After identifying a problem and formulating a plan to address the problem, action is taken. What goes on in the classroom is systematically observed through multiple kinds of data collection procedures, such as audio or video recordings, teachers' diary entries, and observers' notes. Action researchers reflect on the outcome and plan subsequent actions, after which the cycle begins again (Nunan, 1990, 1992).

According to Cohen and Manion (1985, p. 211), action research can also be used to accomplish more specific goals: "(1) to remedy problems in specific situations in order to improve a given set of circumstances; (2) to provide in-service training, giving teachers new skills and greater self-awareness; (3) to inject additional or innovative teaching and learning approaches into a system that normally inhibits change; (4) to improve communication between the practicing

teacher and the academic researcher; and, (5) to provide an alternative to the more subjective, impressionistic approach to problem solving in the classroom.”

McPherson (1997) conducted an action research project in her own ESL class for recent immigrants to Australia. She and 25 other ESL teachers in four states undertook action research projects with students at various levels of English language proficiency. McPherson described three cycles in her action research study. In the first cycle, she reviewed the literature on teaching students with mixed English proficiency levels and experimented with many ways of grouping her students based on their language proficiency. She found that the students appeared to have different goals from hers and sometimes refused to join in the groups and the pairs that she had organized. In the second cycle, she asked the students about the activities and she found that the students were happy to work in mixed proficiency level groups and classes. As a result, she gave more responsibility to the students to select their own materials and activities. As she observed them making their own learning choices, she found that the students had reasons for their choices that she had not anticipated. For example, the students had developed strategies for maintaining civil relations in class, though they had had intragroup tensions because of differing ethnicities and/or the political problems in their home

countries. The teacher's efforts to regroup the students based on their English proficiency levels had inadvertently undermined this delicate balance. Allowing the students more choice was the first step toward resolving this issue. The third cycle was conducted at the end of the course. Although most of the students had begun to work well together, there were two students who were marginalized by the dominant ethnic group of the class. McPherson implemented a strategy of calling on these students and validating their own contributions to the class. As a result, the two students began to become more involved in the class activities.

There are several advantages to action research. First, teachers conduct action research in their classrooms. Second, these projects do not require quantitative data, large numbers of participants, or artificial control over variables. Third, the outcome is applicable to real-world contexts and is likely to improve the efficacy of educational institutions. For example, Tsui (1996) presented a study based on the classroom action research project reports of 38 practicing ESL teachers who were enrolled in the Postgraduate Certificate in Education program at the University of Hong Kong, which is a two-year, part-time in-service secondary school teacher education program. The schools were divided into five bands according to the academic ability of the students. The highest proficiency students were in Band

One while the lowest were in Band Five. The students' English proficiency varied widely, ranging from near-native competence for some upper secondary students in some Band One schools, to students who had difficulty expressing basic ideas in Band Five schools. The action research project involved an examination of the teachers' perceptions of the factors contributing to student reticence, and the documentation of the teachers' attempts to address the problem. In the first cycle, the teachers videotaped or audio-recorded their own lessons and reviewed the tapes in order to identify one specific problem. They then designed a list of strategies to overcome the problem, implemented these strategies for four weeks, and kept a diary of what went on in the lessons for these four weeks. In the second cycle, they videotaped or audio-recorded another lesson at the end of the try-out period and evaluated the effectiveness of their strategies. The strategies were (a) they tried to lengthen the wait time after a question to allow students to think about the question and come up with an answer; (b) some of them tried to improve their question technique by modifying their questions; (c) they informed students that there is not always a 'right' answer and to accept a variety of answers; (d) they allowed students to check their answers with their peers before offering them to the whole class; (e) they provided the students with activities focused on content rather than

form, and; (f) they tried to establish a good relationship with the students. Strategy (a) was not successful in all cases because lengthening the wait time sometimes exacerbated anxiety rather than alleviating it. Strategy (b) was ineffective in that when teachers asked more referential and open-ended questions, some students were put off because the questions generally require long answers. On the other hand, this approach worked more effectively when the students wrote their answers before offering them to the whole class. Strategy (c) encouraged the teachers to be more flexible in regards to students' answers and this attitude encouraged the students to answer their questions. Strategy (d) was successful as some students came to have more confidence in their answers because they had peer support. Strategy (e) was effective because the students were not under the threat of having their mistakes corrected. Several teachers employed strategy (f) and found it effective.

There are also disadvantages to action research. One disadvantage is that action research has not been well accepted until recently in the United States for various reasons (perhaps because of the dominance of the experimental approach), though it has been widely used for many years in Australia, Hong Kong, Europe, and the United Kingdom. The second disadvantage is that relatively few published

examples of action research projects are available in the language classroom research literature in comparison with published examples of other types of studies, and there is still limited professional status associated with conducting action research in some areas. As Markee (1996) pointed out, “the issue of how and where action research is disseminated in fact represents an ongoing problem for advocates of action research” (p. 138). However, in the past decade, several action research studies have been published. The third disadvantage is that at this time no agreed-upon criteria exist for determining the significance of the results of action research, though, in the last two decades, some methodological guidance has been published (Bailey, 2001; Burns, 1998; Nunan, 1990; Wallace, 1998). The findings of action research may not be generalizable because there is only limited control over variables and the participants are not randomly selected from the population. As a result, no strong causal statements are possible. In other words, action researchers usually do not concern themselves with issues of generalizability or causality, because the goals of action research are to develop a local understanding and bring about improvement in a particular context, which means that the results may be limited to an entirely emic perspective.

## A New Approach to Teaching English

As pedagogical theory changes, so do the questions asked of the observed data. To deal with these new questions, data collection procedures are changing in order to provide appropriate material for analysis. In the early 1980s, communicative language teaching (CLT) reached its peak, particularly in North America and Britain. CLT is based on the idea that knowing a language includes more than knowledge of the rules of grammar. Hymes (1970) first proposed a theory of communicative competence and his initial ideas were further developed by other researchers such as Canale and Swain (1980). Efforts were made to empirically validate the proposals made by these researchers (Allen, Bialystok, Cummins, & Mougeon, 1982; Bachman & Palmer, 1981).

Communicative language teaching curricula (Breen & Candlin, 1980; Munby, 1978; Yalden, 1983) and notional syllabuses (Wilkins, 1976) were developed to provide a framework for the communicative needs of L2 learners. The creation of communicative classroom techniques and activities encouraged the more realistic use of the foreign language in the classroom (e.g., Littlewood, 1981). Additionally, proposals were made for an overall methodology of CLT (Brumfit, 1984; Widdowson, 1978). Although CLT was widely accepted and implemented, there

were early indications that it did not mean the same thing to everyone (Johnson, 1982). A number of models and frameworks were proposed—some of which did not include attention to language form while others included attention to both form and meaning (e.g., Allen, 1983; Stern 1983). For example, Allen made a distinction between ‘experimental’ (meaning-based instruction), ‘structural analytic’ (form-based instruction) and ‘functional analytic’ (form- and meaning-based) instruction and suggested how one might incorporate these components to different degrees, depending on the learners’ needs and program expectations. Similarly, Stern introduced an analytic and experimental dimension in which CLT was characterized as moving along a continuum from form-based to meaning-based instruction. Others, however, viewed CLT only as message-oriented practice and argued strongly against the inclusion of form-based instruction within a communicative approach (Newmark & Riebel, 1968; Prabhu, 1979). Little was done to investigate these claims, and the empirical research that was published (e.g., Savignon, 1972) was interpreted by many in favor of meaning-based instruction, even when the outcome of the research indicated that a combination of form and meaning was most beneficial.

Krashen (1985) was one of the strongest advocates of a communicative approach. On the basis of a limited number of second language acquisition (SLA) research findings, he argued that L2 learning was similar to L1 learning and that efforts to create environments in L2 classrooms that were similar to the conditions of L1 acquisition were indispensable. He hypothesized that second language learners acquire the target language if they receive enough comprehensible input and have opportunities to focus on meaning rather than grammatical forms and accuracy.

Other developments in second language acquisition (SLA) research provided further support for a communicative approach. Some researchers (Hatch, 1978; Pica, 1987; Swain, 1985) proposed that L2 learners need to interact in the target language so that they are forced to negotiate meaning. By negotiating what they want to say and what they do not understand, they can arrive at mutual understandings with interlocutors. Through this process, learners were thought to acquire the language forms carrying the meanings they wished to convey.

Communicative language teaching brought radical changes to many L2 classrooms. These changes were described and implemented in general terms, such as comprehensible input and a focus on meaning; however, there were also many

interpretations of CLT and considerable variance in the ways in which teachers implemented it. Under these circumstances, it was necessary for researchers to observe L2 classrooms in order to identify differences in the implementation and effectiveness of these different approaches to CLT. For this purpose, an observation scheme seemed useful.

### *The Need for Process-product Research*

In the 1960s, researchers in the psychometric tradition were heavily product-oriented; as a result, they paid little attention to instructional processes. Because this product-oriented approach led to inconclusive results, some researchers became process-oriented and focused on the description of instructional practices and procedures in L2 classrooms. This change brought with it the use of new research techniques: interaction analysis, discourse analysis, and ethnography. While these process-oriented analyses were being developed, a veritable proliferation of L2 classroom observation schemes appeared in the 1970s and 1980s and considerable efforts were made to systematically observe L2 classrooms. These classroom observation schemes included recording procedures, various types and levels of complexity of evaluation categories, and a focus on a wide range of behaviors.

Some observation schemes were primarily focused on descriptions of pedagogic events, while others were designed to document linguistic behaviors. This research was heavily process-oriented and little attention was paid to learning outcomes.

As researchers needed to investigate both process and product, an observation scheme with well-balanced attention to both process and product was needed. A new scheme, called the Communicative Orientation of Language Teaching Observation scheme (COLT) (Allen, Fröhlich, & Spada, 1984), was designed to help researchers describe classroom processes and examine these processes in relation to learning outcomes (See Appendix A). This observation scheme was also the only one that was based on a theory of communicative language teaching. The COLT is a theory-based instrument (i.e., designed based on CLT principles). I will elaborate more on the theory underlying the COLT at the end of this chapter.

Kumaravadivelu (1999) conceptualized a framework for conducting critical classroom discourse analysis (CCDA). He began with a critique of the scope and method of current models of classroom interaction analysis and classroom discourse analysis, arguing that they offered only a limited and limiting perspective on classroom discourse. Accordingly, he stated that the categories included in an observation scheme reflect the designer's assumptions about what observable

teacher behaviors are necessary in order to build a classroom behavior profile of the teacher and that these principles are clearly reflected in the oldest and the best known scheme in the field of general education, the Flanders Interaction Analysis Categories (Flanders, 1970). He also discussed the COLT observation scheme proposed a decade after the Flanders scheme. While admitting that a significant achievement of the COLT, compared with its predecessors, has been its capacity to help its users differentiate between more and less communicatively oriented instruction, thus enabling them to better connect instructional input with potential learning outcomes, he pointed out some of the limitations of the COLT that characterized other interaction schemes and stated that another method of coding and analyzing classroom data would be more appropriate for a detailed discourse analysis of the conversational interactions between teachers and students.

#### *The Validity of the Observation Categories of the COLT*

The categories included on observation schemes in the past were based on pedagogic rather than psycholinguistic criteria because these schemes were primarily designed for teacher training purposes. As the COLT was based on communicative language theory, it was designed for research in language

classrooms in which some form of CLT is implemented. Because the COLT is theory based, hypotheses about the effect of the theory need to be tested. Moreover, the instructional categories need to be clearly defined in such a way that these hypotheses can be tested in process-product research. Spada, Frölich and Allen (1984) described the rationale, the overall organizational framework, and the specifications of the categories included on the COLT. A year later, Frölich, Spada and Allen (1985) tested the COLT in various programs in order to determine whether the categories could capture the information intended to be observed and if the overall scheme permitted distinctions between more or less communicatively oriented forms of instruction.

In the study mentioned in the previous paragraph, the COLT was used to describe the types of instructions taking place in several French second language programs (FSL), such as core French classes where students received 20-40 minutes a day of FSL instruction; extended French classes where students received instruction in one school subject in French; and French immersion classes in which students received most (or all) of their subject matter instruction in French. The COLT was also used to describe one English second language program (ESL) for school-aged children. Spada et al. (1984) came to the conclusion that the categories

included on the COLT were validated. For example, according to the COLT categories, teachers spent more time focusing on meaning in *more communicatively-oriented classes* and focused more on form and error correction than meaning in *less communicatively-oriented classes*. The researchers were able to differentiate between *more or less communicatively-oriented* classes by observing the teachers' pedagogic focus.

#### *Inter-rater Reliability in the Use of the COLT*

The behaviors described in some of the categories on the COLT are explicit and overt, so these categories are relatively easy to recognize. Examples of explicit categories are *Students participate in their task in a group*, *A teacher teaches in the L2*, and *Students repeat utterance*. Other categories, which deal with more implicit and covert behaviors (e.g., clarification request, and elaboration request), are more difficult to code because observers must use their personal judgment and make a large number of inferences. One way to reduce the amount of subjectivity is to establish interrater reliability by having more than one observer code the same class events independently and verify that they make the same coding decisions for classroom behaviors.

### *A New Version of the COLT*

The COLT is an observation scheme that provides a macroscopic description of L2 classrooms at the level of activity types and the verbal interactions within them.

This instrument was developed in the early 1980's (Allen, Fröhlich, & Spada, 1984; Fröhlich, Spada, & Allen, 1985) and the new version was published in 1995 by Spada and Fröhlich. In the new version, several changes were made in response to specific problems arising directly from the implementation of the scheme and developments in L2 theory, research, and pedagogical issues. These changes led to new definitions of some categories as well as the addition and deletion of others. For example, in the new version, *Text* was added to the *Teacher* subcategory in *CONTEXT CONTROL* in Part A.

### *The Categories on the COLT*

The categories included on the COLT are mostly theoretically driven (See Appendix A). The creators' conceptualization was derived from a comprehensive view of theories of communicative language teaching, theories of communication, and theories of first and second language acquisition research. The COLT has two parts. Part A, which describes classroom practices and procedures at the level of the

activity, is used in real time. Part B, which describes the verbal interactions between teachers and students within activities, is used in post hoc analyses that in most instances are completed after reviewing transcriptions of audio-recorded data. Most of the 73 categories that are distributed across Parts A and B of the COLT represent binary distinctions in instructional practices (e.g., genuine vs. pseudo request; student-centered vs. teacher-centered participation).

As mentioned in the beginning of this section, the categories included on the COLT are mostly theoretically driven. One of the categories, *Participant organization*, in Part A, is related to the belief that group work is essential in the development of communicative competence. In group work, learners are encouraged to try to understand each other by ‘negotiating meaning’. The *Content* category is based on the notion that L2 instruction should be exclusively meaning-oriented and content-based, as these forms of instruction are regarded as communicative because this is the way that children successfully learn their first language. Another category, *Information gap*, in Part B, is based on the idea that in ‘natural’ discourse there is a high degree of unpredictability, so classroom activities and interactions should require the production of answers that are not known in advance. This feature was developed to measure the extent to which instruction

allows for more exchange of unpredictable information in the learning process.

*Sustained speech* is based on the notion that learners benefit from opportunities to ‘stretch’ their linguistic repertoire by engaging in extended course. *Reaction to form or message* is another category that is closely related to the underlying ideas of *Content* in Part A. Research in L1 acquisition shows that caretakers tend to focus on errors in content rather than grammatical errors in children’s speech, and it is believed that this helps the children acquire their first language. This feature was developed to characterize the extent to which teachers and learners react to the meaning of the form of a message. *Incorporation of student/teacher utterances* is based on the theory that teachers’ expanding, developing and elaborating on learners’ utterances contributes to their L2 development. *Form restriction* is another category. In L1 language learning setting, children are thought to formulate a rule system and test hypotheses about how the language works. This is considered to be a crucial component in L1 development. In more traditional L2 classrooms, learners are expected to repeat, imitate, and produce language that is more limited and restricted than naturally occurring language. In communicative language teaching, it is believed that learners’ more creative and uncontrolled language use in the L2 classroom helps them develop their L2. This feature is intended to

measure varying degrees of restriction in terms of linguistic form so that differences along this dimension can be examined in relationship to learning outcomes.

### *The Use of the COLT for Classroom Research*

The COLT has been used in a variety of L2 contexts to examine process and product relationships and to discover matches and mismatches between L2 program goals and practices. For example, two categories of the COLT were reviewed to study the feedback behavior of teachers (Lyster & Ranta, 1997). Fazio and Lyster (1998) compared language learning environments of some classes with the use of the COLT. Rossiter (2003) used the COLT in order to determine what affective strategy instruction might have on learner performance and self-efficacy in speaking tasks. Oliver and Mackey (2003) investigated the role of interactional context in exchanges between teachers and learners in ESL classrooms.

Lyster and Ranta (1997) revised two categories of the COLT in order to study the feedback behavior of four teachers in French immersion classes as they taught science, social studies, mathematics, and language arts lessons to 9- and 10-year old students. The revised categories were *reaction to form/message* and

*incorporation of student utterances.* They also created additional categories that could more fully describe the linguistic behaviors of the teachers and learners in their database so that they could move to a more micro-level description of corrective discourse in L2 classrooms. This research instrument afforded the researchers a more detailed view of L2 classrooms than the COLT. The instrument was developed to describe teachers' reactions to errors and learners' immediate responses (i.e., learner uptake) to this feedback. The model is presented in the form of a flowchart (e.g., *Learner Error* can be dealt with in two ways; *Teacher Feedback* and *Topic Continuation*; *Teacher Feedback* can be responded to in two ways, *Topic Continuation* and *Learner Uptake*). Uptake was defined as "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance" (p. 49).

Fazio and Lyster (1998) compared the second language learning environments of elementary-level students of French in four submersion classrooms<sup>1</sup> and four immersion classrooms in the Montreal area. Their specific

---

<sup>1</sup> Submersion program: a form of bilingual education in which the language of instruction is not the first language of some of the children, but is the first language of others. This type of program is used in many countries where immigrant children enter school and are taught in the language of the host country.

research question was, “What are the similarities and differences in the type of language arts instruction received by L2 learners of French in the submersion context of the French-language schools and the immersion context of the English-language school?” They used Part A of the COLT, which allows for each observed pedagogical activity (the basic unit of analysis) to be coded along five main categories: participant organization, content, content control, student modality, and materials type. Across the immersion classrooms, different observers carried out 29 observations on different occasions. During the observation sessions, the observers checked appropriate categories on the COLT. The results indicated that the approach in the submersion context of the French language schools and that of the French immersion context of the English language schools stood in contrast to one another. The former was strongly analytic and the latter proved to be varied in its integration of analytic and experiential instructional options, including more variety in classroom organization, content that was focused on both language and other topics, and text that included more extended discourse. The researchers did not conclude anything regarding the effects of these approaches because the data in this study were process-oriented, consisting of almost 60 hours of classroom observations that were coded using COLT Part A. These data were supplemented

by transcripts of audio-recordings in the four immersion classrooms and by field notes in the four submersion classrooms. Although the absence of process-product data precludes any conclusions concerning the effects that the two contrasting orientations may have had on L2 learning, the researchers discussed the effects of these programs in terms of social contexts. That is, learners with fewer opportunities for authentic L2 exposure outside the classroom (i.e., immersion students) require more communicative language use in the classroom whereas learners with more opportunities for authentic L2 exposure outside the classroom (i.e., minority-language students) require fewer opportunities for communicative language use in the classroom, yet, minority-language students benefit from pedagogical intervention with an interactive/experiential orientation that integrates a more experimental focus (i.e., journal writing) into their analytic classrooms. They concluded that classroom processes are bound to interact with external variables related to the social context in ways that are worthy of further investigation.

Rossiter (2003) used the COLT in order to determine what effects, if any, affective strategy instruction (e.g., relaxation techniques, listening to music, visualization, and positive self-talk) might have on learner performance and self-

efficacy in speaking tasks. It was a quasi-experimental, non-equivalent comparison-group design, in which one group of adult ESL learners received 12 hours of affective strategy instruction and the second served as a comparison group. Rossiter observed the teachers in the comparison and treatment groups at the beginning of the term and weekly thereafter. He observed 12 hours of the comparison classes, 12 hours of the affective strategy condition with the principal instructor, and nine hours of the affective strategy condition with the substitute teacher who took the place of the principal teacher. The results showed that instruction in affective strategies, such as relaxation techniques, positive self-talk, the use of humor, risk-taking, and self-rewards, provided no significant between-group benefit for L2 performance as measured by speech rate, success, and message abandonment, or perceptions of self-efficacy as measured by task self-efficacy and self-efficacy for learning measured in the narrative task or in the object description task. He argued that the lack of significant between-group differences could be attributed mostly to the fact that the teachers in both conditions strove to develop a sense of community in order to establish a relaxed environment and to encourage the learners to achieve their linguistic goals. These goals were important because most of the participants

were relatively recent refugees to Canada and/or had been out of school for many years.

Oliver and Mackey (2003) investigated the role of interactional context in exchanges between teachers and learners in ESL classrooms. The teacher-learner exchanges were categorized as being primarily focused on context, communication, management, and explicit language. They tried to determine whether distinct interactional contexts can be reliably identified by researchers and teachers in teacher-learner exchanges in classroom discourse and if the opportunity for and the provision and use of feedback differ according to the interactional context. They used a three-part exchange framework: the learner's initial utterance, the teacher's response to the learner's initial utterance, and the learner's reply to the teacher's response. They first identified the learner's non-target-like utterances, then they coded whether the teacher responds to the learner's non-target-like utterances provided negative feedback, and if so, the nature of the feedback they provided. Next, the teacher's feedback was coded based on whether or not their feedback allowed the learners the opportunity to modify their output. Finally, the learner's responses to the teacher's negative feedback were coded based on whether or not they used the feedback by modifying their input. In this study, four unique patterns

of interactional contexts were identified in the data: non-target utterances, feedback provided, opportunities for modified output, and modified output. These context types emerged after discussions of the data based on the COLT. The COLT allowed the researchers to categorize episodes of input and interaction in the L2 language classroom (Spada & Fröhlich, 1995; Spada & Lyster, 1997).

The COLT was designed to help researchers describe classroom processes and examine these processes in relation to learning outcomes. This observation scheme was also the only one that was based on a theory of communicative language teaching. However, the COLT has not been used for pre-service teacher training. The only published example of the COLT being used for training is by Block (1992), who used the COLT in a course called “Classroom Observation” in an MA TESOL program at the University of Barcelona. First, Block showed an excerpt from which she had recorded about half an hour of a colleague’s class. In preparation for the course, she had watched the film with headphones on, and she had seen and heard several phenomena that she had thought would arouse students’ interest. She had intended to let the students watch the film and write anything they thought was noteworthy. However, in the class, Block and the students watched a fuzzy image at low volume so her objectives for the class were not met. Block

believed that the failure was due to a technical problem and came to the conclusion that she would use excerpts from television programs and films. She found their use advantageous because the recordings are of good quality, the materials are readily available, and video allows students to view the excerpts repeatedly.

In the class, the three observation schemes were applied. Block prepared two excerpts from the film *Good Morning Vietnam*. Robin Williams was a teacher focusing on the presentation and practice of expressions that the students would need if they lived on the streets of New York. The students observed the excerpts and analyzed the class with the COLT. Block found that the COLT was the most difficult aspect of the class to explain. Also, in presenting it to the students, she realized that she was opening herself to three criticisms: The first is that she had applied Part A of the COLT, which was designed principally for real-time coding of complete classes, to a 2-minute excerpt from a film. The second is that she had attempted to characterize the degree to which the class was communicative after viewing only a small fragment. The third is that the COLT is an instrument of comparison but she had presented it merely as an evaluative instrument. Regarding the first two points, she argued that the COLT is not easy to handle when people use it for the first time. She also argued that her role was to offer students a way to

begin conducting observations. Her argument suggests that the COLT should be revised depending on how it is used.

To date, no studies have been carried out as far as both teacher training and the use of COLT are concerned. In the first part of the next chapter, I focus on a practicum as part of the teacher licensure program in Japan and observations related with the practicum. I then discuss how observations are conducted for a particular purpose and research conducted in relation to that purpose in the Japanese context. The main purpose is for student-teachers to observe their own teaching on videotape as that should help them analyze their teaching and discuss the feedback provided by their practicum supervisors and college teachers more effectively.

## CHAPTER 3

### TEACHER TRAINING AND OBSERVATION

#### Introduction

One of the primary purposes of classroom observation is to contribute to teacher trainees' professional growth and development. In the previous chapter, the three traditions of classroom research were discussed in terms of products and processes. In this section, classroom research is discussed in the context of teacher education.

According to Leinhardt and Smit (1984), teachers need two kinds of knowledge, subject-matter knowledge and action-system knowledge. Subject-matter knowledge means the specific information needed by teachers to teach content. Action-system knowledge refers to information regarding teaching and learning such as classroom management.

Through the guided, systematic, and focused observation of experienced teachers, student teachers acquire action-system knowledge. According to Day (1990, p. 43), a formal program of observation can help student teachers in four ways: "(a) developing a terminology for understanding and discussing the teaching

process; (b) developing an awareness of the principles and decision making that underlie effective teaching; (c) distinguishing between effective and ineffective classroom practices, and; (d) identifying techniques and practices student teachers can apply to their own teaching.”

There are two broad approaches to observing second language classrooms, qualitative and quantitative. Day (1990) argued that the purpose of the observation must determine the technique and instruments to be used. Qualitative approaches can provide rich, descriptive data about what happens in the second language classroom. For instance, written ethnographies are useful for student teachers in at least two ways. First, the teachers better appreciate the complexities of the second language classroom while attempting to describe what actually happens as classes are being conducted. Second, they become more aware of the multiple roles that second language teachers play. In addition, this approach potentially provides a great deal information about the social context of the classroom, which can be useful in interpreting learners’ behavior.

In addition to the advantages of this approach, there are some disadvantages. The main disadvantage lies in the very nature of qualitative research. It takes a highly trained observer to make competent and reliable observations. As novices,

student teachers can find it difficult to keep up with the rapidly changing sequence of events in the second language classroom. Another disadvantage is that what is perceived is likely to be influenced by the observers' experiences and biases; many observers find it difficult to be completely objective and neutral. A third disadvantage is the large degree of subjectivity that is an inherent part of qualitative approaches.

It is helpful for student teachers if audio and video recordings are made of classroom interactions as they can allow them to observe their own teaching as well as the events taking place in the classroom as a whole. If they initially fail to notice some potentially important events, they can repeatedly listen to and/or watch the recordings and increase the probability of noticing those events. By repeatedly observing the recorded events, student teachers and their teacher mentors can discuss the events that took place and come to an agreement on an interpretation of the event or behavior. In other words, observation supported by audio and/or video recordings is potentially more neutral and objective than observations carried out in real time.

However, this approach is not without problems. The first disadvantage is that recordings can be intrusive if they are not carefully dealt with. Classroom

participants can become nervous in the presence of video cameras, recordings may reveal aspects of the class that are usually not observed, participants might try to behave in ways that they think they are expected to behave, or the presence of recording equipment can hinder participants from paying attention to the class. In addition, not everything can be recorded, and it is also difficult to capture every student as well as the teacher on the video simultaneously. Because most observations in teacher education programs are concerned with the teacher, the most useful results are generally obtained when the camera is focused on the teacher.

Observations based on quantitative approaches generally require the use of a checklist or a form. The instruments used in this approach can be divided into frequency counts or classroom observation scales that are designed to examine teacher behavior, student behavior, or the interaction between a teacher and his/her students or among the students. One advantage of checklists is that even student teachers, who are usually not highly trained observers, can use these instruments because they are relatively easy to understand. Another advantage is that student teachers become better able to discuss teaching after using a checklist. Lortie (1975) pointed out that teachers' abilities to communicate ideas about teaching are

limited if they do not have a common technical language. By using the language provided by observation categories or an observation system, teachers acquire a metalanguage that allows their comprehension and expression of classroom events to become more explicit and analytical, and through the observation and analysis of classroom events, they can more readily identify patterns in the classroom. One further advantage is, as Gebhard, Gaitan, and Oprandy (1990) and others have stated, an observation system helps student teachers focus their observations on particular aspects of the class. This is important because the amount and variety of classroom activity makes it difficult for many student teachers to focus their attention appropriately.

The main disadvantage is that observations are relatively meaningless unless the categories are carefully chosen and their validity established based on theory or previous research. Another disadvantage is that observers tend to focus on the categories listed on the instrument and miss other aspects of the class; the observed behaviors may not explain all of the facts concerning the focus of the observation or the problem.

Categories that require observers to make few inferences are readily recognizable and specific, but those that require more frequent or higher-level

inferences are somewhat unreliable, a key issue in observing any behavior. If an observation system provides a broad set of categories with acceptable validity, it is suitable for student teachers on the condition that the categories requiring student teachers to make high-level inferences are well defined. If more than one observer uses the same system, inter-rater reliability can be established.

A checklist such as the COLT is useful for pre-service teacher training. In this regard, Saville (2002) argued that student teachers as observers should be trained to use the checklists if a reliable and consistent outcome of the observation is to be expected, and observation is one of the essential elements for pre-service teacher training. In sum, an appropriate observation system helps student teachers observe their own teaching and helps supervisors provide useful feedback to student teachers.

### A Practicum for Prospective English Teachers in Japan

In this section, I provide a description of the Japanese teacher licensure system and the Japanese practicum system. There are three levels of the English teacher's certificate for junior high school—special grade, first grade, and second

grade—and two levels of the English teacher’s certificate for senior high school—special grade and first grade.

Students can obtain a second grade certificate for junior high school by completing the teacher licensure program of their junior college, and they can obtain a first grade certificate for junior and/or senior high school by completing the teacher licensure program of their university or a four-year college. They can obtain the special level certificate for senior high school by taking specific courses at the master level. No second grade level certificate for senior high school teachers of English is currently available.

In order to obtain a first grade English teacher’s certificate for junior high school, students are required to earn a minimum of 20 credits from courses in the section related to English, 31 credits from the section related to teaching knowledge and skills, such as psychology and ethics, and 8 additional credits from either section. For the first grade English teacher’s certificate for senior high school, a minimum of 20 credits of courses related to English, 23 credits related to teaching knowledge and skills, and 16 credits from either section are required. The 16 credits are usually related to English.

In most private universities or colleges, the teaching staff is limited. College teachers who are involved in the teacher licensure program frequently teach a teaching methodology course as well as courses such as intercultural communication and English phonetics; however, not every course required for the licensure program is taught by experts in that area.

First-year students who intend to obtain an English teacher's certificate for junior and/or senior high school attend an orientation session in April (at the beginning of the Japanese academic year) where they receive an explanation of the courses they are required to take and also those they can choose as elective courses. The required courses include teaching methodology, a practicum, and the Japanese constitution. Elective courses include English phonetics, English philology, and foreign language acquisition theory.

Most colleges require students to have a certain level of English proficiency before registering for the program. Some colleges specify the level and others do not. In the latter case, students with relatively low English proficiency (e.g., TOEIC < 450) can enroll in the program. When they have successfully passed most of the courses by the end of their third year, they are permitted to participate in a three or four-week practicum.

Before the practicum, the student teachers visit a junior or senior high school where they will have their practicum and attend an orientation session in which the principal gives a lecture on his/her school and its students. They also receive information concerning which class they are required to observe in the first week, and which classes they will teach as student teachers. If they graduated from a junior high school outside of Tokyo, they usually return to teach at that school. If they graduated from a public junior high school in Tokyo, they usually teach at a school located near the junior high school they graduated from. In the case of the practicum conducted at a senior high school, the student teachers usually return to the senior high schools they graduated from.

The practicum takes an average of three weeks to complete and takes place when they are fourth-year university students. The student teachers are required to observe certified teachers' and other student teachers' classes and to teach English. They observe certified teachers' teaching in the first week, and start teaching the following week. On average, they teach between 11 and 20 45- to 50-minute classes (Jimbo et al., 2004). In the case of a two-week practicum, students observe certified teachers' teaching for two or three days of the first week and then start teaching. They teach an average of 6 to 13 classes (Mitsuo & Uchida, 1998;

Takaki, 1997). Regardless of the length of the practicum, at the end of the practicum, the student teacher teaches one class that the headmaster, his/her supervisor, one of his/her college teachers, and other student teachers observe. Student teachers are also required to instruct students in homeroom activities as an assistant to a homeroom teacher, and assist with extracurricular activities (e.g., club activities such as the baseball team).

Student teachers' teaching performance and other requirements are evaluated and the results are sent to their colleges after their completion of the practicum. Evaluation sheets on students' performances are prepared by junior college or university professors. These college evaluation sheets are made up of items that reflect the standards set by the Ministry of Education, Science and Culture (1989), including (a) teaching instruction, which is made up of three subcategories, teaching plans, teaching technique, and evaluation of teaching; (b) student teachers' management of the class and students, and; (c) overall teacher competence, aptitude, and student teachers' working attitude. Practicum supervisors also write comments and give the student teacher a holistic grade that ranges from 1 (the lowest grade) to 5 (the highest grade) (Anonymous, 1993; Kurosaki, 2002; Mitsuo,

2000). Grading criteria are not clear, as a large number of different criteria are used by various supervisors (Anonymous 1993; Kurosaki 2002; Mitsuo, 2000).

After finishing their practicum, the student teachers return to their universities or colleges and ‘reflect’ on the practicum in their teaching methodology course. Reflection requires the student teachers to analyze their practicum based on their supervisors’ feedback, their teaching diaries, and the teaching materials they used during their practicum. They then present the results their analysis in the class, though the time allocated for the presentation can be as short as 5 minutes, and if the class is large, they submit a report to the college teachers in the teacher licensure program rather than make a presentation.

### Feedback

At present, two supervisors serve under the principal when a junior high school or a senior high school accepts a student teacher: the homeroom teacher in charge of a class and an experienced Japanese teacher of English. Student teachers receive feedback as well as advice and suggestions about their teaching and other aspects of their performance from these two supervisors. Experienced Japanese teachers of English provide feedback concerning the student teachers’ teaching

performance. Feedback is usually given through students' teaching journals as well as at a short meeting held after every class and/or when the teachers' daily duties are completed. Common problems with feedback are that it is often subjective and not always specific enough to be of help (Hidaka, 1989).

Kobayashi (1998) investigated how student teachers and college teachers could obtain more specific feedback from practicum supervisors. In the first section of their study, excerpts taken from the teaching practice journals kept by 20 student teachers during their two-week practice period were analyzed in terms of (a) class management, (b) preparation and the appropriate use of teaching materials, (c) establishing rapport with students and acting in a teacher-like manner, and (d) classroom observation. The analysis of the above excerpts revealed the student teachers' personal and professional development. The second section reported the results of questionnaires administered to both the student teachers and their supervisors in order to provide practical information about the student teachers' teaching experience and what they were expected to prepare before their practicum. In the last section, Kobayashi listed ten specific items that guide supervisors' feedback to student teachers. The items included "Is the aim of the class clear?" and "Are the student teacher's explanations clear?" She also suggested that student

teachers who are unfamiliar with classroom observation should be directed to focus on several aspects of the classroom, including their impressions of the class, how the class was conducted, and time allocation. Kobayashi suggested that more specific feedback also be provided.

Nukui, Miura, and Yoshida (2002), who were involved in a teacher licensure program at a Japanese university, attempted to reduce the ambiguity of some supervisors' feedback and allow student teachers to evaluate their own practicum by introducing portfolio assessment. The student teachers recorded what they did in the practicum, and filed the information and other related materials in their portfolios. After the practicum, they reflected on the events that had transpired during their practicum using the file of information. Three problems occurred with this trial. One was that the student teachers did not know how to effectively evaluate their practicum and also did not have time to reflect on the events that occurred while they were involved in the practicum. The second problem concerned cooperation with the supervisors, whose help was necessary to make the portfolio study more successful. The third problem was that most of the supervisors did not know what portfolio assessment was and they had little time to share with the student teachers.

Another way for student teachers to obtain feedback from supervisors is through a meeting after a visiting college teacher observes the student teacher's class with the supervisor, the headmaster, and other student teachers. This is generally the only occasion when student teachers, supervisors, and college teachers can meet together and discuss the student teachers' teaching.

Kizuka (1996) attempted to establish a new way for student teachers to obtain feedback from a meeting with their college teachers and their practicum supervisor. Kizuka's main aim was to improve the quality of the practicum through an examination of the problems that remain to be solved. He emphasized the importance of action research and the establishment of a collaborative relationship among college teachers, student teachers, and supervisors. In this case study, a student teacher's teaching was videotaped. The college teacher, the student teacher's supervisor, and the student teacher observed her teaching by watching the video. Two main problems were identified: her tendency to imitate the supervisor's teaching and to teach the whole content of the textbook exclusively.

The student teacher discussed the problems observed in her teaching with the college professor and her supervisor and she reflected on her teaching. On the basis of the points discussed with them, the student teacher started developing her own

style of teaching. Kizuka stated that most of the problems pointed out in the discussion were resolved as a result of the student teacher's observation of her teaching and that this style of observation was effective in the short-term, pre-service practicum used in Japan. After the action research project was completed, the student teacher's teaching reportedly improved, but the improvement was not clearly described and a number of questions remained unanswered. For example, in what situation was a problem observed? Who was involved? What language was being used to teach? What were the roles of the student teacher and the students?

Kizuka also pointed out two limitations of the approach: its time-consuming nature, and the difficulty of collaboration with supervisors who are busy because they have to supervise student teachers while teaching their own classes and managing extracurricular activities. In sum, although the college teacher and her supervisor suggested ways to look at teaching performance, it was unclear how objective their feedback was.

An optimal way for supervisors to provide feedback to student teachers has not been established yet. Portfolio assessment is too time consuming for many supervisors to implement at present, and Kizuka's action research approach can be misleading if the criteria used are too subjective and the participants are not

familiar with this approach. One possible response to providing useful feedback was suggested by Kaneda (1986), who argued that qualitative and quantitative approaches can be complementary. He stated that both qualitative approaches and quantitative approaches are necessary for classroom research, and qualitative approaches should be based on quantitative ones; otherwise, qualitative studies would be just like talking about researchers' experiences.

Whatever approach is ultimately chosen, it is clear that teacher trainers need to identify an appropriate way for supervisors to provide feedback to student teachers. Moreover, the feedback must be shared among the supervisors, student teachers, and college teachers who are involved in teacher licensure programs, and the approach that is adopted has to be suitable to the English teaching context found in Japanese secondary schools.

#### *Observation Scheme for Providing Feedback to Student Teachers*

In 1989 and 1999, the Japanese Ministry of Education introduced new courses for the study of foreign languages in junior high schools and senior high schools, respectively (Ministry of Education 1989, 1990). One of the primary aims of the new curriculum was to require teachers to focus more on developing students'

speaking and listening skills. In addition, in order to cope with the aims of the new courses of study, licensed teachers are expected to primarily use communicative language teaching (CLT), and student teachers are required to do the same in their practicum. An appropriately revised COLT that can be used by observers of English classes taught by student teachers in Japanese junior high schools is one way to improve their teaching and the practicum system in Japan.

#### Gaps in the Literature

Although classroom observation and evaluation are important issues, few researchers have focused their attention on investigating these issues. This is one of the few studies in which this issue was investigated. Moreover, from the viewpoint of the teaching practicum in Japan, although some studies have been conducted concerning feedback from the practicum supervisor to university professors and student teachers, no researchers have created and used an adapted version of the COLT. The purpose of this study is to investigate a new way of providing feedback to multiple parties involved in the practicum.

## Purpose of the Study and Research Questions

The primary purpose of this study is to suggest a model for a classroom observation scheme that can be used by student teachers, their supervisors, and college teachers when they observe student teachers' teaching and also for student teachers to use when they observe their supervisors' teaching. This model provides a unified and focused viewpoint that enables these three groups to more effectively discuss the student teachers' performance. This model also allows supervisors to provide specific feedback to student teachers who are teaching other classes that the college teachers are unlikely to observe directly. However, these college teachers can later analyze the feedback with the student teachers when their practicum is completed.

The two main goals of this study are to identify (a) a more effective way for student teachers and college teachers to obtain more specific feedback from practicum supervisors, and (b) the causes of differences in the application of criteria used to evaluate student teachers' teaching among student teachers, college teachers, and supervisors. The research questions that are investigated are:

1. What categories should be included on an observation checklist for student teachers, their practicum supervisors, and college teachers to use when

observing student teachers' classroom teaching?

2. To what degree do student teachers, practicum supervisors, and college teachers' ratings of the *Self-assessment Test* differ?
3. To what degree do student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differ when they use the *Revised COLT*?
4. To what degree do student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differ when they use *the Student Teachers' Videotaped Instruction*?

The first research question is discussed in the next chapter.

## **CHAPTER 4**

### **A NEW MODEL OBSERVATION SHEET**

An appropriate observation system is necessary if student teachers are to obtain specific feedback from practicum supervisors and observe experienced teachers' teaching in a focused way. In order to develop a Japanese version of the COLT that is more appropriate to Japanese junior high school classrooms, the revision was implemented in five phases.

When first conceptualized, this study was seen as an exploratory attempt to identify an effective way for student teachers and college teachers to obtain more specific feedback and also to share their criteria with junior high school teachers acting as practicum supervisors. Gradually, the scope of the study widened with the addition of the COLT. In this chapter, how a revised version of the COLT suitable for observations of English teaching at Japanese junior high schools was developed is explained and discussed.

## Phase 1: Revising the Evaluation Sheet

The first phase of this project involved developing a survey of the techniques that teaching supervisors think is relevant to their students. The techniques, which were selected after interviewing three supervisors in 1989 and five in 1990, were added as criteria to the conventional evaluation sheet. The conventional evaluation sheet that supervisors completed at the end of each practicum had criteria a, b, and c (See Table 1); the teaching supervisors wrote comments and awarded grades (1 = lowest, 5 = highest) using these three criteria.

Criteria d, e, f, g, and h (See Table 1) were added to the conventional evaluation sheet. This revised evaluation sheet, *Evaluation Sheet (Version 1)* was sent to supervisors every year between 1990 and 1993.<sup>2</sup> Twenty-three supervisors responded to my request for feedback in those four years.

The eight criteria shown in Table 1 were used to evaluate the student teachers' teaching. The supervisors were asked to write comments and award ratings (1 = lowest, 5 = highest) using the *Conventional criteria* (criteria a, b, and c). They were asked only to award ratings (1 = lowest, 5 = highest) to each of the

---

<sup>2</sup> The *Evaluation Sheet (version 1)* with more criteria *continued* to be sent to supervisors from 1994 to 1996. These added criteria are candidates for the second edition of the evaluation sheet.

Table 1. *Revised Evaluation Sheet Criteria*

Criterion	Explanation of the Criterion
(a) Teaching Instruction	Formulate teaching plans, use teaching techniques, and evaluate the teaching junior high school students.
(b) Class Management	Understand homeroom teacher work, school administration, classroom management, and other school facilities.
(c) Overall Teacher Competence and Aptitude	Attitude towards education, studies on the subject matter, English proficiency, knowledge of English, and originality.
(d) Good Examples	Produce good example sentences to introduce new sentence patterns, provide grammatical explanations, and provide new lexical phrases and vocabulary.
(e) Model Reading	Provide a good model reading.
(f) Accuracy	Communicate in English with junior high school students accurately and appropriately and correct students' grammatical mistakes.
(g) Communicative Language Activity	Provide students with communicative language activities.
(h) Communication with Assistant Language Teachers	Communicate with ALTs in order to prepare for team teaching classes and give a demonstration in front of students in class.

*Note.* ALT = Assistant Language Teacher.

*Classroom English proficiency criteria* (criteria d, e, f, g, h) when they thought that they had observed the student teachers' teaching sufficiently. Each supervisor observed the student teacher several times before awarding the ratings.

Descriptive statistics for the eight criteria are shown in Table 2. The means for the *Classroom English Proficiency Criteria* (criteria d, e, f, g, h) were much lower than the *Conventional Criteria* (criteria a, b, and c); the supervisors did not award high

scores to the student teachers for English proficiency and teaching effectiveness. This was useful information, given that the student teachers' English proficiency should be reasonably high if they are to meet the guidelines set by the Ministry of Education (1989), which emphasize the use of communicative language teaching. Notwithstanding weaknesses still present in the *Evaluation Sheet (Version 1)*, it allowed the supervisors to provide more specific feedback to the student teachers than the conventional evaluation sheet.

Table 2. *Revised Evaluation Sheet Criteria Descriptive Statistics*

Criterion	<i>M</i>	<i>SD</i>
(a) Teaching Instruction	4.74	0.45
(b) Class Management	4.61	0.45
(c) Overall Teacher Competence and Aptitude	4.61	0.50
(d) Good Examples	3.04	0.56
(e) Model Reading	3.09	0.52
(f) Accuracy	3.09	0.52
(g) Communicative Language Activity	3.00	0.52
(h) Communication with Assistant Language Teachers	3.04	0.64

The ratings from the supervisors also agree with the results of the survey administered by Ibe (1993), in which supervisors were asked what was lacking in university and college English language programs in Japan. The majority of the respondents stated that universities and colleges should provide their students with

programs in which the students can improve their communicative competence in English. Students' practicum supervisors stated that the student teachers should improve their communicative competence in English, especially their speaking skills, before participating in a practicum, as this would allow them to be more effectively trained to teach more communicatively during their practicum and also to be ready to teach English professionally after graduation.

There are two other possible interpretations of the results displayed in Table 2. One is that the *Conventional criteria* are primarily concerned with teaching; their scope covers issues such as the teaching plan and the evaluation of each class, and as these skills are largely unrelated to foreign language proficiency, the means of the *Conventional criteria* are relatively high. In contrast, the *Classroom English proficiency criteria* are more concerned with the student teachers' English proficiency and the effectiveness of their performance in English, skills that are quite difficult to acquire. For this reason, the means of the last five questions are relatively low.

The second interpretation is that the scores of the *Conventional criteria*, which were originally on the conventional evaluation sheet, are influential when new teachers are selected. Although paper examinations and interview tests play

important roles in hiring decisions, what the conventional sheet says about a candidate is also taken into consideration. As a result, many supervisors avoid awarding low ratings because they can cause student teachers not to be selected as full-time teachers, even after passing several demanding tests. The *Classroom English proficiency criteria* are not referred to, because they are not widely used yet.

One serious limitation with the *Evaluation Sheet (Version 1)* concerns rater leniency and severity. The conventional evaluation sheet lacks a clear criterion by which grades are awarded, so student teachers with better English proficiency and better performances during the practicum are sometimes given lower grades than less proficient student teachers teaching at a different junior high school because of differences in the severity of different supervisors. *Evaluation Sheet (Version 1)* also faces the same problem.

Another problem with *Evaluation Sheet (Version 1)* is that the *Conventional criteria* from the conventional evaluation sheet represented broad categories that partially include the *Classroom English proficiency criteria*, which are partly subcategories of two of the *Conventional criteria*, *Teaching Instruction* and *Overall*

*Teacher Competence and Aptitude*, so the range that the *Classroom English proficiency criteria* covers is already partly covered by those two criteria.

There were two main findings in this phase of the study. The first was that *Evaluation Sheet (Version 1)* provides college teachers and student teachers with useful feedback from the practicum supervisors because they can be relatively specific and instructive. The second finding was that the *Evaluation Sheet (Version 1)* should be systematic so that college teachers and student teachers understand the problems that student teachers' teaching supervisors observe during the practicum.

#### Phase 2: Developing Additional Criteria

In 1994, 1995, and 1996, I interviewed the 17 supervisors, observed five student teachers' teaching during their practicum, and observed their supervisors' teaching in order to collect a new list of criteria for a new evaluation sheet. In the interviews, I asked these 17 supervisors to look at the list of criteria on the revised evaluation sheet (See Table 1), delete any criteria they thought unnecessary and/or add any they believed were missing. I collected a new list of criteria through the interviews and through my observation of the five student teachers' teaching and the five supervisors' teaching (See Table 3).

The new evaluation sheet, *Evaluation sheet (Version 2)*, which has 12 criteria, differs from the *Evaluation Sheet (Version 1)* in three ways. First, the *Conventional criteria (Teaching Instruction, Class Management and Overall Teacher Competence and Aptitude)* on the *Evaluation Sheet (Version 1)* are excluded from the *second version*.

Table 3. *A New List of Assessment Criteria*

- 
- a. Instruct students in English.
  - b. Make an ‘oral introduction’ of a new sentence pattern.
  - c. Provide good examples of the new structure.
  - d. Explain new structures well.
  - e. Answer questions concerning grammar.
  - f. Introduce context in your own words in English.
  - g. Ask questions in English about the text and have students answer in English.
  - h. Do a good model reading.
  - i. Translate English texts into Japanese if necessary.
  - j. Plan an authentic and interesting language activity.
  - k. Notice students’ serious English mistakes and correct them immediately.
  - l. Answer questions about the Japanese meaning of an English word.
- 

Second, in comparison with the *Evaluation Sheet (Version 1)*, *Version 2* includes more criteria concerning if and how well student teachers encourage students to practice the four skills. This newly revised list on *Version 2* is more communicatively oriented and student-centered while the *Classroom English proficiency criteria* on the *Version 1* is more focused on the student teachers’

English proficiency. Although the student teachers' English proficiency is an important determination of their teaching effectiveness, their proficiency should not be the main focus of attention; instead, their English use needs to be viewed in relation to their teaching and the students' learning.

Third, criterion (h) in Table 1, which concerns how well student teachers can communicate with ALTs when preparing for their team teaching class and giving a demonstration to the students, was deleted from the list shown in Table 3. This does not mean that the student teachers did not have to team teach; however, they have few opportunities to team teach because the ALTs come to each junior high school only a few times a month and they visit any particular class a maximum of once a month. As a result, ALTs often come to the classes that the student teachers teach before and after the practicum but not during it.

One problematic issue was pointed out while the survey was being designed: there is no standard by which supervisors can evaluate student teachers' performances; thus, each supervisor evaluates his/her student teachers' performance using his/her own standards, sometimes in consultation with one or more teachers working at the same school. As a result, the scores that the student

teachers receive for their practicum performance strongly depend on the supervisors' standards and degree of severity.

In the following phase, *Evaluation Sheet (Version 2)* is discussed further. The first purpose is to confirm the validity of the twelve criteria on *Evaluation Sheet (Version 2)* with the student teachers. The second purpose is to add and/or delete criteria on *Evaluation Sheet (Version 2)* based on feedback from the student teachers.

### Phase Three: Student Teachers' Self-evaluation

As discussed in the previous phase, I made a new list of criteria to be used on the *Evaluation Sheet (Version 2)* from 1994 to 1996. The intention was to create a systematic evaluation sheet, an observation scheme, which could be used by supervisors to provide feedback to student teachers and by student teachers to self-evaluate their teaching. In order to make a systematic evaluation sheet, I needed to confirm the validity of *Evaluation Sheet (Version 2)* and revise it if necessary. In order to accomplish this, I conducted two studies: The first involved investigating student teachers' practicum in 1996 and the second involved looking at the criteria on *Evaluation Sheet (Version 2)* from theoretical perspectives.

## Investigating Student Teachers' Practicum in 1996

Twelve Kita-Saitama college students in the teacher licensure program participated in this study. The participants had just finished their two-week practicum teaching English at junior high schools as student teachers in June, 1996. Each student teacher taught 12 or 13 class periods for first-, second-, and third-year junior high school students and about four periods for each level. Data were collected with a questionnaire, interviews with the student teachers, an inspection of the student teachers' teaching diaries, observations of their teaching on videotape, and feedback from their supervisors.

By the end of June 1996, the participants returned to the college from the practicum and continued their university studies. All of them seemed to have more confidence than before the practicum. In the first class of a teaching practice course after the practicum, I asked the participants to complete *The Student Teachers' Practicum Teaching Questionnaire* (See Appendix B). Twelve of the 16 students agreed to complete the questionnaire.

First, I explained the meaning of each question and asked them to take their time and ask questions if they had any. It took them approximately one hour to complete the questionnaire. Afterwards, they submitted their student teacher

diaries, their teaching plan for every class they had taught, and videotapes on which one period of their teaching was recorded. I also asked them to answer Questions 7 and 8 on the questionnaire again in the following class. The participants were permitted to change their answers if they wished. The test-retest reliability coefficient for Questions 7 and 8 was acceptable at .94.

I observed the participants' teaching in the practicum on videotape, read their teaching plans for the classes I observed on the videotape, and analyzed their participant diary entries. While I was engaged in the above work, I talked with some of the participants about their teaching shown on the videotape.

In the third class of the teaching practice course, I interviewed the 12 participants in order to gather more details about their teaching plans, their diary entries, and their answers to *The Student Teachers' Practicum Teaching Questionnaire*, and to determine whether their answers corresponded with the answers they had previously provided (See Appendix C for the analysis of their answers for Questions 3, 4, 5, and 6, which concern their English proficiency.)

The participants were asked to describe their teaching procedure (Questions 1 and 2 in Appendix B) for the following reason. In questions 7 and 8 on *The Student Teachers' Practicum Teaching Questionnaire*, they were asked to check the criteria

that were important when teaching English in junior high school because they needed to evaluate their own teaching performance using the criteria on the questionnaire. The left side of Table 4 shows the 14 criteria used for Questions 7 and 8 on *The Student Teachers' Practicum Teaching Questionnaire*. The right side of the Table shows the 12 criteria on the *Evaluation Sheet (Version 2)*. The 14 criteria in Questions 7 and 8 on the *Student Teachers' Practicum Teaching Questionnaire* consist of the 12 criteria on *The Evaluation Sheet (Version 2)* plus two more criteria, (1) and (2), shown on the left side of Table 4.

It was necessary to verify how many of these criteria their teaching procedure included in order to determine why some of the questions were not answered and whether any inaccurate responses had been given.

Question 7 asked the participants to select one of four responses (1 = *not important*; 4 = *very important*) for each of 14 criteria.<sup>3</sup> Question 8 asked the student teachers to evaluate how well they were able to use each item in the practicum.

---

<sup>3</sup> In the last section, the 12 criteria are introduced. Two more criteria were added to *To do a good model reading* in order to determine to what degree the participants found English pronunciation difficult. These additions resulted in a total of 14 criteria on the questionnaire. The order of the criteria introduced in the last section was changed on later versions of the questionnaire.

Table 4. *Abstract from the Student Teachers' Practicum Teaching Questionnaire*

<i>Abstract from the Student Teachers' Practicum Teaching Questionnaire</i>	12 criteria on the <i>Evaluation Sheet (Version 2)</i>
1. Pronounce individual words accurately.	
2. Pronounce at the sentence level accurately.	
3. Instruct students in English.	a. Instruct students in English.
4. Provide good examples of new structures.	c. Provide good examples of new structures.
5. Explain new structures well.	d. Explain new structures well.
6. Answer questions about the Japanese meaning of an English word.	l. Answer questions about the Japanese meaning of an English word.
7. Answer questions concerning grammar.	e. Answer questions concerning grammar.
8. Translate English texts into Japanese if necessary.	i. Translate English texts into Japanese if necessary.
9. Make an 'oral introduction' of a new sentence pattern.	b. Make an 'oral introduction' of a new sentence pattern.
10. Do a good model reading.	h. Do a good model reading.
11. Ask questions in English about the text and have students answer in English.	g. Ask questions in English about the text and have students answer in English.
12. Introduce context in your own words in English.	f. Introduce context in your own words in English.
13. Notice students' serious English mistakes and correct them immediately.	k. Notice students' serious English mistakes and correct them immediately.
14. Plan an authentic and interesting language activity.	j. Plan an authentic and interesting language activity.

In Table 5, row A1 shows the means of the fourteen items. Higher numbers indicate greater perceived importance. Row A2 shows the mean scores awarded by

the student teachers based on how well they thought they had taught while they observed their own teaching on the videotapes.

Eight of the student teachers teaching first- or second-year students chose 1 or 2 for Criterion 8 (*Translate English texts into Japanese if necessary*), perhaps because translation is usually not used in these classes. Four participants chose 3; they all taught third-year classes in which translation occupies more class time. For self-evaluation (See section A2 in Table 5), most of the participants chose 3 or 4 for Criterion 8 (*Translate English texts into Japanese if necessary*). They were accustomed to translating from English into Japanese, and some of the student teachers teaching the third-year students spent more time translating than those teaching the first- and second-year students.

Table 5. *The Student Teachers' Self-evaluation of their Teaching Performance*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A1	4.0	3.8	3.9	3.3	3.4	3.5	3.5	2.4	3.3	3.8	3.5	3.0	3.6	3.0
A2	2.3	2.3	1.8	2.0	2.5	3.0	2.0	3.0	2.0	2.3	1.9	1.6	2.5	1.8

Only one participant awarded herself more than 40 points for her own teaching performance. The videotape of this participant confirmed that she was teaching well in regards to the 14 criteria, and this participant's supervisor gave her

42 points. The participant with the lowest self-assessment, 16 points, was not teaching well in relation to the criteria, and the participant's supervisor gave her 15 points. Given the similarity of their scores, the participants and their supervisors appeared to have a similar understanding of the criteria.

#### A Theoretical Perspective of the Evaluation Sheet (Version 2) Criteria

In the previous sections, the 14 criteria were listed and their validity was confirmed. However, before attempting to create an observation scheme, I wanted to confirm (a) that these 14 criteria were placed in the same order that they are likely to appear in the teaching procedure (e.g., the oral introduction of the text usually occurs early in the lesson) and (b) how these criteria related to each other in the teaching procedures. My discussion of the criteria is based partly on theoretical perspectives drawn from the second language acquisition literature. In order to accomplish this, I discuss one class period of each college student's teaching. The period was selected by asking the college students which period they thought they had best planned and taught.

*Textbooks Authorized by the Ministry of Education and Classroom Instruction*

Each student teacher taught an average of 13 50-minute class periods during the 1996 program. Seven of the 17 student teachers taught seventh graders, nine taught eighth graders, and seven taught ninth graders (See Appendix D).

Junior high school English textbooks, which are based on the guidelines set by the Ministry of Education in 1989, are published for each grade. Each textbook consists of approximately 10 to 15 units, and each unit has three to five subunits. In this study, subunits are numbered as follows: In *Sunshine English Course 2, Program 4-2*, Program 4 is the 4th unit of the textbook and 2 is the second subunit of the fourth unit.

One reading text is divided into several passages, each of which is introduced with a new sentence pattern and vocabulary. Whether one passage is taught in one period (50 minutes) or two periods depends mainly on the degree of text difficulty, and how the new sentence pattern is introduced and practiced. Junior high school English teachers are supposed to cover the entire textbook within one academic year, so there is an element of time pressure in the curriculum. When one period is spent covering one passage of the textbook (one period type), the teacher usually begins with a greeting, a review of the previous class, an introduction of new

materials (e.g., a new grammatical structure), language activities such as an information gap activity, reading (e.g., an oral introduction of the text), and consolidation, in which the student teachers sum up their lessons, assign homework, and/or ask questions regarding what they have taught in the lesson. When two periods are taught together (two period type), one period is made up of an oral introduction of the new materials and language activities as well as greetings and a review of the previous class. The second period is made up of reading and other activities focused on areas such as writing, grammar, and vocabulary.

### *Teaching Procedures*

In order to clarify how the student teacher taught, I divided their teaching lessons into pre-activities, main activities, which include reading and interaction activities (e.g., information gap activities), and post-activities. The lines between these divisions are sometimes blurred because the notion of pre/while/post-reading activities is not always clearly implemented in the 17 examples of the student teachers' teaching (See Appendix E).

### *Pre-activities*

In the 17 examples, all the lessons except lesson G included the introduction of new vocabulary. Except for lessons I and Q, all the student teachers introduced new lexis by having the students practice the pronunciation of the new words, and by translating them into Japanese. This approach shows an exclusive focus on preparing the readers for likely linguistic difficulties in the reading text.

In lesson I, the student teacher explained summertime in London, and helped her students to understand the difference between summer in London and in Japan in order to better comprehend the meaning of text. The word *summer* was familiar to the students, but explaining it served to activate existing schema (Wallace, 1992). In lesson Q, the student teacher explained *had better*, but only in terms of its grammatical usage, not its sociolinguistic or pragmatic usage.

An oral introduction of the reading text is also a common activity that was used by nine out of the 17 student teachers. In the oral introduction, the student teachers outlined the text orally in English in order to help the learners understand the text when they were allowed to open their textbooks and read. Many of the student teachers and supervisors believed that they could "give" students content

schema rather than activate what the learners already knew in the oral introduction of new reading materials.

Generally speaking, the "Oral introduction" approach proposed by H. E. Palmer (1921) has been widely used at junior high schools throughout Japan. In a discussion of this issue, Sano (1995) argued that the aim of the oral introduction is not just to introduce the content of the passage, but also to lead students to activate their world knowledge and make inferences. He also stated that some Japanese junior high school English teachers use the "oral introduction" to introduce conceptually difficult words that are related to what students are familiar with, to provide background information about the reading text, and to activate the students' schema.

In lesson N, the student teacher distributed Japanese reading materials about World War II and the A-bomb in Hiroshima. The students were going on their school trip to Hiroshima the following autumn, and these materials were part of the school curriculum. The Japanese materials covered much of the content of the text, and the student teacher provided an outline of the text in simple English as a pre-reading activity. From the viewpoint of interactive reading processes, these two

activities are similar to the oral introduction adopted by some other student teachers.

Fujimori (1995) provided an example of "oral introduction" and other pre-reading activities. She first wrote the word *war* on the blackboard and asked the students to write words associated with the word. Then the students presented the words that they had written and she wrote those words on the blackboard with the word *war* in the center. She then connected the student-generated words with the word *war* by drawing lines between them. She then started an oral introduction designed to activate the students' schema and allow them to take a stance on the text when they read.

In lessons C and J, the pre-reading activities encouraged interactive reading processing. Important concepts in *The Time Gap* were explained beforehand, and some background knowledge was provided in lesson C. In lesson J, the student teacher tried to activate the students' schema by asking several questions while pointing at some pictures. This technique has been shown to be effective (Omaggio, 1979). In lesson P, the only activity involved the students in identifying a new sentence pattern in the text, and was thus focused on new and/or difficult expressions.

A true/false reading comprehension test was administered in five of the student teacher lessons. The purpose of the test was to assess whether the students had understood the listening text and also to use the questions as a way to focus the students' attention on specific information in the listening passage.

#### *Main Activities Including Reading and Interactive Activities*

Reading aloud tasks were adopted in all the lessons (See Table 6). Makita (1996) stated that students should practice reading aloud to express their comprehension of the text. Mori (2003), referring to Perfetti (1985), wrote that reading aloud occurs in L1 reading if readers read expressions that they are unfamiliar with or when they are not fluent readers; he suggested that the same phenomena are likely to occur in L2 reading and that reading aloud in the L2 is a step to fluent reading.

Translation was adopted in six of the student-teacher lessons; however, if the four lessons that ended with the distribution of a written translation to the students are counted, translation was used in ten lessons. Difficult grammatical points were explained in six classes so that the students would better comprehend the reading

texts. In four of the classes, explanations of new grammatical structures were provided in the passages that were translated.

Table 6. *While-reading Activities in the 17 Student-Teacher Lessons*

	A1	B1	C1	D1	E1	F1	G1	H2	I2	J2	K2	L2	M2	N3	O3	P3	Q3
R	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+
T				+		+			+						+	+	+
G									+		+	+					+
O		+	+		+			+	+	+			+				
t								+			+	+		+			

*Note.* R = reading aloud practice; T = translation; G = grammar explanation; O = other activities; t = translation sheet provided at the end of the class.

One main teaching skill taught by the student teachers is reading; however, the amount of reading is closely related to the students' grade level (Ibe, 1993), with students in higher grade levels reading more. Many Japanese junior high school teachers say that listening and speaking are emphasized in the seventh grade, and reading becomes increasingly important as the students move to higher grades, because the textbooks for students in higher grades have more complicated sentence structures, more vocabulary, and involve a wider variety of topics. Ninth grade students also need to prepare for senior high school entrance examinations, which are primarily focused on reading and grammar.

Eight student teachers' lessons included activities other than translation:

Asking questions about the text (J, M), practicing frequently used expressions for communication on the phone that were part of the text (B), administering a true-false reading test (C, D), and providing background knowledge about a reading passage (H). If students can read flexibly, actively, and reflectively by finding answers to questions about the text, asking questions is successful, but the questions asked in C, D, J, and M only required the students to read part of the text. In lesson E, the students needed to identify which picture was being introduced while reading the text.

In lesson I, the procedure appeared to lead the students to read interactively. For pre-reading activities, the student teacher provided explanations of some words in the text, gave the students time to read silently, asked them some questions about the text, and then asked them to translate sentences that were the answers to the questions the students had difficulty answering. She urged them to try a bottom-up reading strategy in order to improve their comprehension by having them translate word by word. Her questions involved the students in guessing the meaning of one new word that she had not explained, identifying the names of sight-seeing spots in London, and finding words related to the picture on the page.

### *Post-activities*

The four main activities observed in the student teacher lessons showed the students engaging in role-plays, memorizing the reading text that the class had covered in that period, taking part in speaking practice using expressions in the text, and completing a translation task.

Though some classes did not include post-reading activities, it is possible that I might have included some post-reading activities in the while-reading activity category. For example, memorizing the reading text covered in that period and completing the translation task could be considered post-reading activities in some cases.

In the previous sections, I have discussed several aspects of the student teachers' reading classes based on the data I gathered from the student teachers' diaries, interviews with the student teachers, and the junior high school teachers' feedback in the student teachers' diaries. Several criteria were added to the 12 criteria on the *Evaluation Sheet (version 2)* based on the discussions regarding (a) activating students' background knowledge as part of pre-reading activities; (b) having students read a text individually; (c) having students read a text orally in

unison; (d) memorizing a text; completing a dictation drill; and (e) translating Japanese sentences into English.

Table 7. *The List of Criteria (1996 Version)*

---

a. Instruct students in English.
b. Make an 'oral introduction' of a new sentence pattern.
c. Provide clear examples of a new structure.
d. Explain a new structure well.
e. Answer questions about English grammar.
f. Activate students' background knowledge as part of pre-reading activities. Teachers may apply the "Oral introduction" (Palmer, 1921) approach to activate students' background knowledge.
g. Introduce context in your own words in English.
h. Ask questions in English about the text and let students answer in English.
i. Do a good model reading.
j. Read a text orally together.
k. Let students read a text individually.
l. Translate English texts into Japanese if necessary.
m. Plan an authentic and interesting language activity.
n. Notice students' serious English mistakes and correct them immediately.
o. Answer when a student asks about the Japanese meaning of an English word.
p. Memorize part of a reading text; present a dictation drill.
q. Translate a Japanese sentence into English.

---

At this point, the list of candidate criteria for an observation scheme consisted of the 12 criteria on the *Evaluation Sheet (version 2)* plus the five criteria listed in the previous paragraph (Table 7). However, two categories of activities are mixed: One category appears as the class proceeds, and the other category can appear throughout the class. Examples of the former category are greeting, introduction of new materials, and consolidation, and examples of the second category are a

teacher's use of the first language/the second language of the students, instruction, and students' modality. Through the phases I took, I came to realize that the observation scheme I intend to make must be able to record what activities of the latter category were occurring while an activity from the first category was being taught. Otherwise, it would not be possible to provide sufficiently precise and accurate feedback. The activities in the first category were listed clearly enough for the student teachers, college teachers, and junior high school teachers to check, but the activities in the second category were not sufficiently clear. Activities belonging to the second category, which were based on CLT, needed to be added to a new observation scheme. This led me to revise the COLT, which is made up of numerous activities in the second category.

### Revision of the COLT for the Student Teacher Practicum at Japanese Junior High Schools

Two goals were addressed by revising the COLT in order to make it more appropriate for use in the educational contexts investigated in this study. The first goal was to identify a more effective way for student teachers and college teachers to obtain more specific feedback from practicum supervisors. The second goal was

to revise the COLT so that student teachers, college teachers, and practicum supervisors can use it more effectively when observing, analyzing, and evaluating the student teachers' teaching.

The following six procedures were followed when revising the COLT:

1. I observed five experienced junior high school teachers' English classes. These teachers were the student teachers' practicum supervisors.
2. The 17 criteria (See Table 7) were used to check the five teachers' instruction.
3. I videotaped the classes.
4. I noted the activities that were involved in each of the 17 criteria.
5. I revised the COLT based on steps 1 to 4 above.
6. I observed two other experienced teachers' English classes in order to pilot the revised COLT.

The revised COLT is for student teachers to use in their practicum when they assess their own teaching by observing their videotaped lesson and for their supervisors and college teachers to use when they observe the student teachers' teaching and discuss the student teachers' teaching performance with them. College teachers observe the student teachers' class once during their practicum and it is the

only opportunity for these three people to discuss the lesson. Junior high schools set aside about one hour for this discussion; thus, the revised COLT needs to be simple enough to be used easily during the discussion while also covering important aspects of the teaching situation.

### *The Structure of the COLT*

The COLT observation scheme is divided into two parts (See Appendix A). Part A concerns classroom events at the level of episode and activity, and Part B is focused on the communicative features of verbal exchanges between teachers and students and/or students and their peers as they occur within each episode or activity. Of the 73 categories included on the COLT, most represent binary distinctions in instructional practices (e.g., genuine vs. pseudo requests; student-centered vs. teacher-centered participation).

Along the top of the table in Part A of the COLT are categories such as *Time*, *Activities and episodes*, *Participant organization*, and *Content*, and several of the categories have subcategories. For example, *Class*, *Group*, and *Individual* are subcategories of *Participant organization*. Some subcategories, such as *Class*, *Group*, and *Individual* are further subcategorized. For instance, *Class* is further

divided into *Teacher to student or class, student to student, or Student to class, and Choral work by students.*

### *Time and Activities and Episodes*

These categories have no subcategories. Under the category *Time*, observers fill in the time when one activity or episode begins. Under the category *Activities and episodes*, observers fill in the name of an activity or episode that they observe. Usually there is more than one activity or episode in one class, so multiple spaces are provided for this category.

### *Target Language, Information Gap, Sustained Speech, and Reaction to Form or Message*

Along the top of Part B are categories concerning the use of *Target language, Information gap, Sustained speech, and Reaction to form or message.* These categories are further subcategorized. For example, *Target language* has two subcategories: *L1* (Use of the native language) and *L2* (Use of the second/target language). *Information gap* has two subcategories, *Giving information* and *Requesting information*, and some of these subcategories are further divided. For

example, *Giving information* is divided into *Predictable* and *Unpredictable*, and *Requesting information* is divided into *Pseudo requests for information* and *Genuine requests for information*. Part B is used in post hoc analyses that in most instances are completed after reviewing transcriptions of audio-recorded data.

### The Revised COLT

As mentioned in the previous section, *Time* and *Activities and episodes* have no subcategories. Under the category *Time*, observers note the time when one activity or episode begins. Under the category, *Activities and episodes*, a relatively large space is provided for observers to note the name of the activity or episode that they observe. Most of the 17 criteria I have made in the last several years fall into the *Activities and episodes* category.

First, I sorted the 17 criteria into two groups: One group was placed in *Activities and episodes* and the other was placed in other categories of the COLT. Two criteria, (a) *Instruct students in English* and (n) *Notice students' serious English mistakes and correct them immediately*, occur frequently, so they were added as major categories at the top of the revised COLT (see Table 9). The remaining 15 criteria were added to *Activities and episodes*.

After observing the five experienced teachers' teaching, I added *Review* to the revised COLT because I observed the teachers conducting reviews of the previous lesson (See Table 8). I also divided criterion (m), *To plan an authentic and interesting language activity*, into two sections (*Students' interactive activities* and *Evaluation*), each of which involves interactions between the teacher and students and/or students and their peers. I also combined four criteria into one category, *Model/Chorus reading/etc* in the Revised COLT: (h) *Ask questions in English about the text and let students try to answer in English*; (i) *Do a good model reading*; (j) *Read a text orally together*; (k) *Let students read a text individually*, because criterion (h) is not always present and takes little time to complete when it is present. I could not observe any interactions in the activities of the other three criteria. I observed the student teachers asking students if they had any questions, so I added *Qs and As* and combined the following two sub-criteria into *Qs and As*: (e) *Answer questions about English grammar* (o) *Answer when a student asks about the Japanese meaning of an English word*.

I simplified the *Activities and episodes* category. The order in which these activities and episodes appear in the student teachers' lessons is almost perfectly consistent across the five student teachers' classes, and the order is familiar to the

Table 8. *Activities and Episodes on the Revised COLT*

Criteria	<i>Activities and Episodes on the Revised COLT</i>
	Review
b. Make an ‘oral introduction’ of a new sentence pattern	New sentence pattern
c. Give good examples of the new structure; (d) Explain a new structure well	Grammatical explanation; Repetition
	Introduction of an interactive activity
	Explanation of the game; Demonstration
m. Plan an authentic and interesting language activity;	Students’ interactive activities
	Evaluation
g. Introduce context in your words in English	Introduction of new words
f. Activate students’ background knowledge as part of pre-reading activities	Oral introduction of the text
e. Answer questions about English grammar; (o) Answer when a student asks about the Japanese meaning of an English word	Qs and As
h. Ask questions in English about the text and have students answer in English; (i) Do a good model reading; (j) Read a text orally together; (k) Let students read a text individually.	Model/Chorus reading/etc.
p. Memorize a text; Dictation drill	Recitation of text
l. Translate English texts into Japanese if necessary	Translation into Japanese
q. Translate a Japanese sentence in English.	Part of consolidation

student teachers and their supervisors. The name *Activities and episodes* was changed to *Procedures of the class*, as this label is more familiar to the student teachers and junior high school teachers. Many of the names of the criteria and activities in the category *Procedures of the class* were shortened in order to simplify the COLT table (See the right hand column in Table 8).

#### *Other Categories on the Revised COLT*

I revised the original COLT on the basis of the observation and memos that I had made while observing the five experienced teachers' teaching English at junior high schools. These same teachers were the student teachers' supervisors. At that time, I found that the activities in some of the categories in Parts A and B were not evident in the five teachers' classes and the other two experienced teachers' teaching, so I deleted one unnecessary category from Part A and one from Part B. One of the main categories in Part A is *Materials*, in which there are two subcategories: *Type* and *Source*. *Source* is usually a textbook authorized by the Ministry of Education, so I deleted this subcategory. In Part B, there are two main categories: *Teacher verbal interaction* and *Student verbal interaction*. In the latter category, there is a subcategory, *Sustained speech*, whose subcategories are *Ultra-*

*minimal, Minimal, and Sustained speech*. I deleted *Sustained speech* because I had not observed any sustained speech occurring in junior high school classes.

Revisions of the categories were also carried out so that I was able to unite these two parts into one.

It is also necessary to explain some categories listed in Table 8. As discussed above, the COLT is theory-driven, as it is based on CLT, so the categories included on the COLT are mostly theoretically driven. However, the Revised COLT was revised partly based on the basis of the observation and memos that I had made while observing the five experienced teachers' teaching English at junior high schools, so there are categories such as *Grammatical explanation; Repetition, and Translation into Japanese* that are not entirely theory-driven or based on CLT. In other words, the revised COLT is based on a Japanese version of CLT and not every aspect of it is theoretically driven.

I divided *Procedures of the class* into five blocks because five student teachers appeared on the video. These five blocks were labeled *Block A, Block B, Block C, Block D, and Block E*. *Review and New sentence pattern* are in *Block A*. *Grammatical explanation, Repetition, and Introduction of interactive activities* are in *Block B*. *Explanation of the game; Demonstration, Students' interactive*

*activities*, and *Evaluation* are in *Block C*. *Introduction of new words*, *Oral introduction of the text*, *Qs and As*, and *Model/Chorus reading/etc* are in *Block D*, and *Recitation of text*, *Translation into Japanese* and *Part of consolidation* are in *Block E*.

There are common categories across *Procedures of the class* that are divided into five blocks. Besides these common categories, there are categories particular to certain *activities* and *episodes* in one or more than one block. Below, I explain these common categories and the categories particular to certain activities in *Procedures of the class*.

### *Common Categories*

One common category is *Topic* and *Teaching materials*. Another is *focus of the lesson*, which has four subcategories:

1. *grammar, vocabulary, and/or pronunciation at the word level and/or sentence level;*
2. *function;*
3. *discourse and/or sociolinguistics, and*
4. *culture.*

*Teacher's activities* is also divided into four subcategories:

1. *Use of language (L1/L2);*
2. *Interlocutor (individual/class or group);*
3. *Purpose of use of the language (content/socializing or instruct), and*
4. *Teacher's questions (pseudo/genuine).*

*Students' activities* is divided into two subcategories:

1. *Use of language (L1/L2);*
2. *Student modality (listening, speaking, reading, writing).*

#### *Categories Particular to Certain Activities*

Some categories are particular to certain activities in the *Procedures of the class* section in one or more than one block. As I observed the seven experienced teachers' teaching, it became clear that some categories were needed in order to observe certain activities in the teaching procedures section, while other categories were unnecessary. For instance, one subcategory of *Student activity*, *Interlocutor (teacher/peers)*, is necessary across *Blocks A, B, C* and *D*, but it is not necessary in *Block E*. A second example is that another subcategory of *Student activities*,

*Sustained speech (ultra-minimal/minimal)* is necessary in *Blocks A, B and C*, but not in the other blocks.

### The Revised COLT and its Categories

The revised COLT has five main categories (See Table 9): *Time, Procedures of the class, Topic, Teaching materials, Focus of the lesson, Teacher's activities,* and *Student activities*. *Time, Procedures of the class,* and *Topic* have no subcategories. The starting time of each activity of *Procedures of the class* is noted in the *Time* category. *Procedures of the class* includes the 15 criteria or activities listed above. Observers note the topics of the classes they observe in the *Topic* category.

### *Focus of the Lesson*

*Focus of the lesson* has four subcategories:

1. *Grammar, vocabulary, and/or pronunciation at the word level and/or sentence level;*
2. *Function;*

Table 9. *The Seven Main Categories and Subcategories on the Revised COLT*

Main categories	Subcategories	
<i>Time</i>	No subcategory	
<i>Procedures of the class</i>	15 criteria or activities that have already been explained previously (See Table 8.)	
<i>Topic</i>	No subcategory	
<i>Teaching materials</i>	No subcategory	
<i>Focus of the lesson</i>	<i>Grammar, vocabulary, and/or pronunciation at the word level and/ or sentence level</i>	
	<i>Function</i>	
	<i>Discourse and/or sociolinguistics</i>	
	<i>Culture</i>	
<i>Teacher's activities</i>	<i>Interlocutor</i>	<i>Class or group</i>
		<i>Individual</i>
	<i>Use of language</i>	<i>L1</i>
		<i>L2</i>
	<i>Purpose of use of the language</i>	<i>Content</i>
		<i>Socializing or instruct</i>
	<i>Teacher's questions</i>	<i>Pseudo</i>
		<i>Genuine</i>
	<i>Sustained speech</i>	<i>Ultra-minimal</i>
		<i>Minimal</i>
<i>Incorporation of student/teacher utterances</i>	<i>Form</i>	
	<i>Repetition</i>	
	<i>Correction/paraphrase</i>	
	<i>Comment on content</i>	
	<i>Clarification request</i>	
	<i>Elaboration request</i>	
<i>Student activities</i>	<i>Use of language</i>	<i>L1</i>
		<i>L2</i>
	<i>Interlocutor</i>	<i>Peers</i>
		<i>Teacher</i>
	<i>Sustained speech</i>	<i>Ultra-minimal</i>
		<i>Minimal</i>
	<i>Student modality</i>	<i>Listening</i>
		<i>Speaking</i>
		<i>Reading</i>
	<i>Participant organization</i>	<i>Writing</i>
<i>Individual</i>		
<i>Class</i>		
<i>Students' Interactive activities</i>	<i>Class/group/pair</i>	
	<i>Sentence pattern (fixed/not fixed)</i>	
	<i>Vocabulary (fixed/words to choose/free)</i>	
	<i>Interaction (free/to some extent free/fixed)</i>	
		<i>Sustained speech (ultra-minimal minimal)</i>

3. *Discourse and/or sociolinguistics*, and;

4. *Culture*

“At the word level and/or sentence level” was added to the original definition of the *Grammar, vocabulary, and/or pronunciation at the word level and/or sentence level* category because grammar, vocabulary, and pronunciation are usually dealt with at the word level and/or the sentence level in the Japanese junior high school classes that I have observed. *Function* is a reference to functions/communicative acts (e.g., requesting, apologizing, and explaining).

*Discourse and/or sociolinguistics* refers to the way in which spoken or written sentences combine into cohesive and coherent sequences such as when describing a process (e.g., how to plant a herb garden) and/or a reference to spoken or written forms or styles appropriate to particular contexts (e.g., the difference in the use of *may* and *can* in formal and informal contexts). *Culture* is a reference to the cultural background of words, lexical phrases, and the text used in the class.

### *Teacher's Activities*

*Teacher's activities* has five subcategories: *Interlocutor*, *Use of language*, *Purpose of use of the language*, *Teacher's questions*, *Sustained speech*, and *Incorporation of student/teacher utterances*. Each of these subcategories is further sub-subcategorized.

*Interlocutor* refers to the person(s) that the speaker or speakers are talking to, and it is further subcategorized into *class or group* and *individual*. *Class or group* indicates that the teacher is speaking to the whole class or one or more than one group. *Individual* indicates that the teacher is speaking to an individual.

*Use of language* has two sub-subcategories, *L1* and *L2*. *L1* is the use of the students' L1 (Japanese) and *L2* is use of the L2 (English). *Purpose of use of the language* is subcategorized into *Content* and *Socializing or instruct*. *Content* indicates that the teacher is speaking to teach and *Socializing or instruct* indicates that the teacher is giving greetings or instructing students.

*Teacher's questions* is subcategorized into *pseudo* and *genuine*. *Pseudo* indicates that the speaker possesses the information requested (e.g., 'Who is the author of the book that we are reading today?') and *genuine* means that the

information requested is not known in advance by the questioner (e.g., ‘Where did you go last Sunday?’).

*Sustained speech* is intended to measure the extent to which speakers engage in extended discourse or restrict their utterances to a minimal length of one sentence, clause, or word. It is subcategorized into *ultra-minimal* and *minimal*. *Ultra-minimal* means that student turns consist of one word only or two-word speech fragments, such as an article and a noun (e.g., *the sun*) and a preposition plus a noun (e.g., *at home*). *Minimal* indicates that the teacher-student turn consists of more than one or two words, long phrases, or one or two main clauses or sentences. The original COLT has one other sub-subcategory, *sustained*, which indicates that a teacher-student turn consists of at least three main clauses. As all of the teacher talk by the student teachers I observed was either *ultra-minimal* or *minimal*, I did not include the category of *sustained speech*.

*Incorporation of student/teacher utterances* has six sub-subcategories: *form*, *repetition*, *correction/paraphrase*, *comment on content*, *clarification request*, and *elaboration request*. *Form* is a reaction to the linguistic form (e.g., grammar, vocabulary, and pronunciation) of the preceding utterance(s). *Repetition* is full or partial repetition of previous utterance(s). *Correction paraphrase* is any linguistic

correction of a previous utterance or indication of incorrectness and *paraphrase* is reformulation of previous utterance(s). When a teacher's reaction to the form of the students' utterances is observed, he corrects or paraphrases them. *Comment on content* is positive or negative response (not correction) to previous utterance(s). Comments can be either message-related or form-related. *Clarification request* is requests indicating that the preceding utterance was not clearly understood and a repetition or reformulation is required. *Elaboration request* indicates a request for further information related to the subject matter of the preceding utterance(s). Requests for an explanation are included in this category.

### *Student Activities*

*Student activities* is divided into six subcategories: *Use of language*, *Interlocutor*, *Sustained speech*, *Student modality*, *Participant organization*, and *Students' interactive activities*.

*Use of language* concerns the use of the *L1* (Japanese) and *L2* (English).

*Interlocutor* refers the person(s) to whom a speaker or speakers are talking to and it is subcategorized into *peers* and *teacher*. *Sustained speech* has two sub-subcategories: *ultra-minimal* and *minimal*. *Ultra-minimal* means that students

turns which consist of one word only or two-word speech fragments such as article plus noun and preposition plus noun. *Minimal* means that the teacher and a student turns which consists of more than one or two words, long phrases, or one or two main clauses or sentences. *Student modality* is a reference to which skill students are engaging in: *listening, speaking, reading, and writing.*

*Participant organization* refers to the way in which students are organized. Three basic organizational patterns make up this category: *individual, class, and class group/pair.* *Individual* means that students work on their own either on the same task or the different tasks. *Class* indicates (a) a central activity led by the teacher in which the teacher interacts with the whole class and/or with individual students; (b) a central activity led by a student or students (e.g., a group of students acts out a skit and the rest of the class is the audience); (c) the whole class or individual groups participate in choral work (e.g., repeating a model provided by the textbook or teacher). *Class/group/pair* means that one or several activities are explained by the teacher and practiced by the whole class and each group or pair then works on the same task or different tasks.

*Students' interactive activities* was added for observing teaching at a junior high school. In many of the classes I have observed, teachers let students practice

their English through games such as information-gap activities (e.g., interview games). Games consist of three elements: the targeted sentence pattern, vocabulary, and interaction; thus, I included four subcategories in the *Students' interactive activities* category: *Sentence pattern (fixed/not fixed)*, *Vocabulary (fixed/words to choose/free)*, *Interaction (free/to some extent free/fixed)*, and *Sustained speech (ultra-minimal/minimal)*.

#### Rationale for and Further Explanation of the Categories

##### *Time and Procedures of the Class*

Each activity and episode is timed so that a calculation of the percentage of time spent on the various features of the revised COLT can be determined. All other activities are coded within the context of each activity of Procedures of the class.

##### *Topic*

Topic refers to topics related to the students' immediate environment and experiences (e.g., family and community topics) or topics that extend beyond the

classroom and immediate environment (e.g., international events). Topics are determined by the textbooks authorized by the Ministry of Education and Science.

### *Teaching Materials*

Teaching materials are generally decided as teachers are supposed to use textbooks authorized by the Ministry of Education and Science. Therefore, the important point from the viewpoint of communicative teaching and learning is how they are used in this context. The way that the textbooks are used is reflected in how English is practiced in the class.

### *Use of Language in Teacher's Activities, Student Activities, and Other Categories*

This category is based on the assumption that the amount that the students use the target language is closely related to their L2 development. In many approaches to second language acquisition, input is seen as an important factor underlying acquisition, though the place of input within those approaches varies. For example, from the viewpoint of the Input Hypothesis, input needs to be comprehensible (Krashen, 1985). This category is closely related with *Purpose of use of the language (content or instruction/socializing) in the Teacher's activities*

category. This category permits an investigation of whether more communicative interactions tend to take place in the L1 rather than the L2 in classrooms in which the teacher and the students share the same L1.

#### *Teacher's Questions in Teacher's Activities*

Interaction in English occurs most often between the teacher and student(s) in Japanese junior high schools. It starts with the teacher's question, which can be display questions (*pseudo*) or referential questions (*genuine*). In the communicative language teaching paradigm, information gap activities are one of the main types of communicative task, so classroom activities in which the answers are not known in advance are frequently used. This feature was developed to measure the extent to which teacher questions allow for giving and receiving unpredictable information and creating more interactions between the teacher and students.

#### *Incorporation of Student/Teacher Utterances in Teacher's Activities*

L2 language acquisition researchers have argued that building on the L2 learners' previous utterances can contribute to language acquisition. This feature is intended

to measure the different ways in which this may be carried out. This category, as mentioned above, has six subcategories: *form*, *repetition*, *correction/paraphrase*, *comment on content*, *clarification request* and *elaboration request*.

Negotiation of meaning is assumed to contribute to L2 acquisition (Gass, 2003; Long, 1996). In situations where negotiation of meaning takes place, *clarification request* and/or *elaboration request* can occur. If second language speakers understand that the listeners do not understand them and they are in a situation where they have to make themselves understood as clearly as possible, they need to express themselves again by modifying what they have said.

Corrective recasts are also believed to be important for L2 acquisition (Doughty & Varela, 1998; Farrar, 1992; Han, 2002) and can be checked in the *correction/paraphrase* category.

At the same time, the *incorporation of student/teacher utterances* category was developed to investigate whether the teacher's reaction to students' utterances is form-focused (*form*) or meaning-focused (*comment on content*).

### *Sustained Speech in Teacher's Activities and Student Activities*

Sustained speech is not often observable in Japanese junior high school English classes. If it takes place, it is usually in response to an elaboration request from the teacher. If the teacher asks a student to elaborate, and the student responds, they can achieve minimal sustained speech. I have yet to observe sustained speech consisting of at least three main clauses between the teacher and a student, or between students.

Requests for repetition and clarification sometimes occur when students are engaged in interactive activities; thus, sustained speech is possible between students.

### *Student Modality in Student Activities*

Traditionally, each language skill has been taught in isolation in L2 instruction in Japan, so learners often engage in listening activities separately from speaking activities. Students are encouraged to integrate the four skills to reflect a more authentic use of language in communicative language teaching classrooms. This category identifies the various skills involved in a classroom activity.

*Participant Organization and Students' Interactive Activities in Student Activities*

In the literature on communicative language teaching, group work is considered indispensable for learners to develop communicative competence. For instance, Swain (1995, p. 128) stated: “Output may stimulate learners to move from the semantic, open-ended, non-deterministic, strategic processing prevalent in comprehension to the complete grammatical processing needed for accurate production. Output, thus, would seem to have a potentially significant role in the development of syntax and morphology” (p. 128). In addition, it is through production that learners are able to receive implicit or explicit feedback, engage in hypothesis testing, and develop automaticity (Gass, 1997; Swain, 1995). This contrasts with teacher-centered classes in which learners spend more time responding to the teacher’s questions and rarely initiate discourse. Participant organization is intended to describe distinctions between teacher-centered and group-work interactions in L2 classrooms. How often and to whom the teacher and the students communicate is also reflected in the *Interlocutor in Teacher’s activities* and *Student activities* categories.

In this section, I have discussed how I have developed the revised version of the COLT used in this study. In the next section, I introduce a pilot study using the revised COLT and related questionnaires.

### The Pilot Study

The pilot study was conducted in June 2000 in order to (a) investigate the reliabilities of The *Teaching Skill Questionnaire* criteria, *the Revised COLT*, and The *Student Teachers' Videotaped Instruction*, (b) determine whether the participants would understand *the Revised COLT*, and The *Student Teachers' Videotaped Instruction* (See Appendix H & I), (c) investigate the clarity of the teaching extracts on the 60-minute videotape on the basis of which the *Student Teachers' Videotaped Instruction* were answered and *the Revised COLT* was used, and (d) validate the *Revised COLT*. The contents of the *Student Teachers' Videotaped Instruction* and the *Teaching Skill Questionnaire* (See Appendix G) as well as the videotape were based on the teaching activities that I had made in the previous several years. The items had been checked through several phases. This pilot study was also designed to investigate the possible presence of the Rashomon

Effect,<sup>4</sup> confirm the clarity of the 60-minute videotape and also see that the participants would understand the activities recorded on the videotape. If the Rashomon Effect was operating, different participants would interpret the same scene differently and the answers would vary. This would be indicated by the participants' choosing wrong activities.

### *Participants*

The participants were five female student teachers, two male and three female college teachers, and five junior high school teachers (practicum supervisors), four of whom were male and one of whom was female. The five student teachers had been chosen from the four-year college (Hachioji College)<sup>5</sup> where the researcher taught as a full-time instructor. These five student teachers were chosen randomly from those who had completed the main questionnaire.

---

<sup>4</sup> Rashomon Effect: The Rashomon effect is the effect of the subjectivity of perception on recollection, by which observers of an event are able substantially different but equally plausible accounts of it. It is named for Akira Kurosawa's film *Rashomon*, in which a crime witnessed by four individuals is described in four mutually contradictory ways. The film is in turn based on two short stories by Akutagawa Ryunosuke, though the technique is not as manifest in the prose as in the film.

<sup>5</sup> The names of the educational institutions mentioned in this study are pseudonyms.

One of the five college teachers was a colleague involved in the teacher licensure program, another used to be in an Ed.D. cohort at the same graduate school I attended, another taught in the teacher licensure program of Iruma University located in Saitama prefecture, and the last two taught in the teacher licensure program at a college in Ishikawa prefecture. One of the five junior high school teachers was an English teacher at a Tokyo public junior high school for 20 years, and the other four taught at a private junior and senior high school attached to a women's university located in central Tokyo. All of the participants signed written consent forms stating that the data could be used in this study.

### *Materials*

Five kinds of materials were used in the pilot study: a contract, a 10-minute videotape that was accompanied by a set of questions, the *Teaching Skill Questionnaire*, the *Revised COLT*, and the *Student Teachers' Videotaped Instruction* and that the participants answered after observing the five student teachers' teaching on the 60 minute-videotape.

The 10-minute videotape consisted of 10 excerpts, each of which was between 30 to 45 seconds in length (See Appendix F). In each excerpt, one of the

five student teachers, who appeared in the 60-minute videotape, could be observed teaching in their practicum. Ten activities were chosen for use in the 10 excerpts from typical teaching activities. The five student teachers appear twice consecutively in the same order as they would in the 60-minute videotape. The participants were asked to choose which activity one of the student teachers was engaged in and to record their responses on the answer sheet.

The *Teaching Skill Questionnaire* (Appendix G) consisted of 15 questions that were based on typical teaching activities. Each question was answered on a 7-point Likert scale (1 = *a participant finds it very easy to teach*; 7 = *the most difficult to teach*). These 15 questions were ordered in terms of hypothesized difficulty, so the first question was likely to be answered with 1 or 2, the questions in the middle with 3 or 4, and the last question with 6 or 7. The five student teachers had been asked previously to arrange these 15 items in order of difficulty, and the order of the questions in *The Teaching Skill Questionnaire* was determined based on the answers provided by the five student teachers.

The participants watched the five-section 50-minute videotape that included intervals where I explained how to answer part of the *Revised COLT* and *the Student Teachers' Videotaped Instruction* for the next section. The videotape

showed English teaching from reviews to consolidation with the five student-teachers. The participants used the *Revised COLT* (See Appendix H) while observing the teaching.

All the criteria were based on my past findings concerning the COLT. All the criteria except the greeting activity were adapted into the left side of the *Revised COLT* table. Thus, the categories on the left side showed how the class would proceed and what teaching activities were likely to be observed. The participants' roles were listed at the top of the table. The participants checked what was prominent in the top categories for each activity listed in the teaching procedures. If one category was particularly prominent, the participants were asked to mark the category with a double circle, and if prominent, they would mark it with a single circle.

The *Student Teachers' Videotaped Instruction* consisted of 42 5-point Likert scale questions (See Appendix I). The participants were asked to respond to the questions based on what they observed on the videotape. Most of the questions were paired. One was a general and theoretical question and the other asked how well the student teacher performed a particular activity in each section of the video. These questions were developed based on the 15 items.

### *Procedures*

I first explained the research purposes and the pilot studies to the five student teachers in July, 2000. I then asked them to sign the consent form (See Appendix J). After obtaining their written consent, I gave each person a set of questionnaires with the videotape. They answered the *Teaching Skill Questionnaire*, watched the 10-minute videotape with the *Revised COLT* and *The Student Teachers' Videotaped Instruction*, and then began watching the 60-minute videotape. In the 60-minute videotape, I first introduced the components of the videotape, explained how and what to mark on *The Revised COLT*, and how to answer the *Student Teachers' Videotaped Instruction*. The participants then watched the videotape. Some of the participants needed a few minutes to answer the remaining questions. After approximately 5 minutes, the answer sheets were collected.

On the same day, I sent a set of the questionnaires with a videotape to the five college teachers and the five practicum supervisors. I had previously told them about the research and the pilot studies by phone or in person. Within a month, I was able to collect all of the answer sheets.

## Results

### *The Teaching Skill Questionnaire*

A one-way ANOVA and a Scheffé post hoc comparison were conducted to identify differences in the means of the answers of the three groups of participants to each question on The Teaching Skill Questionnaire. A Bonferroni adjustment was applied and the  $p$  level was set at .003 as 15 comparisons within the same data set were made. Only question 15 showed a statistically significant difference,  $F(2, 12) = 14.35, p < .001$ .

Table 10. *Comparison of the Answers of the Teaching Skill Questionnaire*

*Among the Three Groups*

Source of variance	SS	df	MS	F	P
Between Groups	38.93	2	19.47	14.35	.001
Within Groups	16.40	12	1.37		

Though it was found that the responses to only one item differed significantly, the differences in the means of the other answers of the three groups were quite large. The Cronbach alpha reliability estimate was .88, so the reliability of the self-assessment instrument was acceptable.

### *10-minute Video*

The responses were analyzed using Cohen's kappa coefficient (N\*K matrix) (Seigal & Castellan, 1988). The result was  $K = .764$ , and the significance of  $K (z) = 8.179$ . As the scenes the participants watched were excerpts from the 60-minute videotape, the reliability of the participants' responses to the 60 minute-videotape was confirmed.

### *60-minute Video*

The participants drew a circle or a double circle in the appropriate box on the Revised COLT when they perceived a particular activity on the videotape. For example, if a participant thought that a student teacher was engaged in a review activity, the box describing the activity of 'speaking in English as a teacher' placed on the top of the table and also the activity titled *reviews* on the left side of the table was checked with a circle or a double circle. The score for each box was calculated; a circle was counted as one point and a double circle was counted as two points.

All three groups awarded scores for the same activities. The student teacher group and the college teacher group awarded high scores to some activities to which the JT group gave no scores or only one score. For example, in the use of

English by the student teachers on the 60-minute videotape, the student teacher group and the college teacher group perceived the student teacher using English and yet the junior high school teacher group rarely perceived this behavior. As the groups have different backgrounds and their experiences concerning teaching English at the high school level vary, it is unsurprising that their perception differed.

A one-way ANOVA and a Scheffé post hoc comparison were conducted to identify possible differences in the means of the answers of the three groups to each question. A Bonferroni adjustment was determined using <http://www.quantitativeskills.com/sisa/calculations/bonfer.htm>; the  $p$  level was set at .001. No question was statistically significant, and yet the differences of the means of the answers were not small enough to be ignored. As the number of the participants is much larger in the main survey, it is likely that some variables will show statistically significant  $F$  values, given the greater statistical power that comes with larger  $N$ -sizes. Cronbach's alpha estimate of reliability was .80.

## **CHAPTER 5**

### **METHODS**

#### Participants

Three groups of participants took part in the study: student teachers, college teachers, and junior high school teachers. The student teacher group was made up of 57 Japanese females who were approximately 22 years of age. They were studying at a four-year women's college and were in their fourth and final year at the school. They had already finished a two- or three-week teaching practicum at a Japanese junior or senior high school. After completing their practicum, they returned to the college and continued attending an English teaching methodology class that was offered as part of the teacher licensure program. In the class, I explained about the study and asked the students to participate. They were informed that participation was voluntary and that participation (or lack of participation) would not affect their grades; 57 of the 61 students agreed to participate, and they signed the written consent form shown in Appendix H.

The college teacher group was made up of 19 Japanese university professors, all of whom were involved in a teacher licensure program for junior and senior

high school teachers of English. Two were my colleagues and seven were people I knew as teachers in teacher licensure programs at other universities. I asked the other ten participants to take part in the study by e-mail after obtaining their names from the JALT (Japan Association of Language Teachers) members' list. This group of 19 professors was made up of 12 male and 7 female professors. I made a list of these participants, and after calling them on the telephone and explaining the purposes of the study, I sent the questionnaires with a videotape and a written consent form to those who agreed to participate.

The third group of participants was 28 junior high school teachers (12 Japanese male and 16 Japanese female teachers). I knew 20 of the teachers through the practicum that my college students participated in. I asked these 20 teachers to introduce other junior high school teachers of English who had experience supervising student teachers. I made a list of the candidates in this group, and after calling them, explaining the study, and receiving their verbal consent to participate, I sent the questionnaires with a videotape and a written consent form that I asked them to sign (See Appendix J).

## Instrumentation

Four instruments were used in the main study: a written consent form, a questionnaire about 15 teaching skills, a 60-minute videotape with a checklist, and a 42-item questionnaire. All participants permitted me to use the data that I obtained from the questionnaire by signing the consent form. As the participants in the pilot study had little trouble answering the questions after watching the videotape, the same materials were used in the main study.

### *The Teaching Skill Questionnaire*

The *Teaching Skill Questionnaire* was made up of 15 questions based on the 15 items concerning teaching English in junior high schools (See Appendix G). The purpose of the instrument was twofold. The first purpose was to determine whether there was a significant difference between the three groups of participants for any of the teaching techniques. The second purpose was to determine the correlation between their responses to the other questionnaires and their responses to the *Teaching Skill Questionnaire*. The respondents were asked to judge how easy or difficult each teaching item was on a 7-point Likert scale (1 = *It is very easy to teach*; 7 = *It is very difficult to teach*). The 15 questions were ordered in terms of

their expected endorsability; thus, the first question was expected to be answered with a rating of 1 or 2, the questions in the middle with 3, 4, or 5, and the last question with 6 or 7. In order to arrive at an order of hypothesized endorsability, the five student teachers participating in the pilot study were asked to arrange the 15 items in order of perceived difficulty; this order was then used on the *Teaching Skill Questionnaire*.

#### *The Revised COLT*

The *Revised COLT* was designed to provide observers with a common viewpoint when observing, discussing, and evaluating the student-teachers' teaching so that the student teachers can obtain specific and systematic feedback about their teaching performance and the junior high school teachers can provide more systematic evaluations of the student teachers' performances in the practicum.

The participants used the *Revised COLT* (Appendix H) to evaluate the five student teachers' English teaching performances on the five-section videotape. A different student teacher appears in each of the five sections. The purpose was to identify significant differences between the ratings of the student teacher group, the college teacher group, and the junior high school teacher group.

The *Revised COLT*, which was an adapted version of the original COLT, was based on past findings concerning the COLT (see Chapter 4). The left side of the COLT indicates how the class proceeds, while information concerning the teacher's and learners' roles is listed at the top of the table. The users of the instrument check what is prominent in the top categories for each teaching procedure. In the case of this study, the participants were asked to mark a prominent category with a single circle, and an especially prominent category with a double circle.

### *The Student Teachers*

In this section I describe the five student teachers. I knew four of the student teachers very well as I taught them at the college. I did not know very much about the fifth student teacher, as she was recommended to participate in the study by a member of a professional teacher organization in Japan (Japan Association of College English Teachers (JACET)) who taught her at a different college.

Two of the student teachers on the video were familiar with junior high school students, as they taught at a prep school for junior high school students. The other student teachers knew junior high school students socially because they had brothers and/or sisters who were junior high school students; as a result, they knew

their siblings' junior high school friends. At least one of the student teachers eventually became a part-time junior high school teacher, and another continued to teach at a prep school after graduation. Four of them wanted to be teachers, so they sat teacher certification examination. All of them failed, but one initially became a part-time teacher, and five years later was able to become a full-time teacher.

All of the student teachers were hard working and they made considerable efforts to succeed in their practicum. Student teacher A was a calm, but very determined person. Student teacher B, who was from another college, was outspoken and had an honest attitude. The teacher who recommended that she take part in the study stated that she was cut out to be a teacher. Student teacher C has a gentle, cheerful, personality. Student teacher D was a free and easy type. Student teacher E was a little older than the others, as she had worked for two years before entering the college.

#### *The Student Teachers' Videotaped Instruction*

The *Student Teachers' Videotaped Instruction* was made up of 42 multiple-choice questions measured on a 5-point Likert scale (Appendix I). The 42 questions were made in accordance with the 60-minute videotape, which has five segments: Video

A, Video B, Video C, Video D, and Video E. As a result, there were five groups of questions, each of which was closely related to a particular video segment. The participants answered the questions based on what they saw on the videotape. Most of the questions were paired: One question from each pair was a general, evaluative question (e.g., Is it useful to make students aware of the new sentence pattern and new part of the English grammar while reading?) and the other concerned how well the student teacher was teaching from this viewpoint (e.g., The student teacher was trying to make students aware of the new sentence pattern and new English grammar while reading. Was she teaching effectively?). These questions, which were written based on the 15 items discussed in Chapter 4 (See Table 7), were used to determine on what specific points the three groups of participants differed.

### Procedures

I asked the student teacher participants to come to the classroom on a certain day. I counted the number of participants, called the roll, handed out a set of questionnaires to each participant, and asked them to provide their written consent to participate. Next, they completed the *Teaching Skill Questionnaire*. This took approximately 15 minutes. I then briefly explained the contents of the other two

questionnaires (The *Revised COLT* and the *Student Teachers' Videotaped Instruction*) and directed them to mark the checklist on the *Revised COLT* with a circle or a double circle while watching the videotape. The participants then started watching the five videotape segments. I answered questions after they had finished watching the videotape and had completed the questionnaire. I then collected the consent forms, the *Teaching Skill Questionnaire*, and the questionnaires with the checklist, which were not separated from the answer sheets.

On approximately the same day, I sent a set of questionnaires to the college teachers and junior high school teachers who had agreed to participate in the study. I included a letter saying that I would be very appreciative if they could complete the questionnaires and send back them to me within six months. I also included a stamped self-addressed envelop. Finally, I checked each set of documents after they arrived and I entered the responses into an EXCEL file.

### Analyses

Four research questions are investigated in this study. The first research question is: What categories should be included on an observation checklist for student teachers, their supervisors, and college teachers to use when observing

student teachers' classroom teaching? This research question was discussed in Chapter 4. The COLT was revised based on the results.

The second research question is: To what degree do student teachers, practicum supervisors, and college teachers' ratings on the *Teaching Skill Questionnaire* differ? The data from this research question were used to identify patterns in the responses from the three groups to the 15 items on the *Teaching Skill Questionnaire*.

The third research question asked to what degree the student teachers, practicum supervisors, and college teachers' ratings of the student teachers' teaching performances differed when they used the *Revised COLT*. In order to answer this question, I focused on the ratings that the participants awarded the student teachers' *L2 Use* and four other related categories as demonstrated on the video. These five categories are some of the categories on the Revised COLT (See Table 9). The raters used a partial credit rating scale of 0, 1, or 2 according to how much the student teachers used English as their second language in the class (L2 use) and related categories (activities), such as the purpose of using the L2 (socializing or content), the types of questions they asked (pseudo or genuine), the

students' use of English, and sustained speech between student teachers and students (ultra-minimal or minimal).

I selected one teaching situation section from each video in order to simplify the analysis and to permit the participants to observe each student teacher's L2 use. The video sections were selected by determining sections where the student teachers had rich opportunities to use English. Block A (the first student teacher, *New sentence pattern*), Block B (the second student teacher, *Introduction of interactive activities*), Block C (the third student teacher, *Evaluation*), Block D (the fourth student teacher, *Model/Chorus reading*), and Block E (the fifth student, *Part of consolidation*) were selected. There were about 200 blocks, so in order to reduce the raters' burden, I filled in some blocks with a circle, double circle, or a cross; the cross indicated that the block did not have to be completed. The following blocks were marked with a cross: L2 use in Block E; Purpose of the use of the language (socializing) in Block C, and Sustained speech (ultra-minimal) in Block D. These data were analyzed with the Rasch rating scale model and the ratings that each group give to the five student teachers were compared and analyzed with a one-way ANOVA.

The fourth research question asked to what degree student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differed when they used the *Student Teachers' Video-taped Instruction*. Because all the items on the *Student Teachers' Video-taped Instruction* were measured with a common rating scale, the Rasch rating scale model was used to analyze these data. The questions were divided into three groups according to the wording used on each 5-point Likert scale. One Likert scale measured the student teachers' performance from 1 (*Not useful*) to 5 (*Very useful*). The second Likert scale measured the items on a scale from 1 (*Not effective*) to 6 (*Very effective*), and the final Likert scale group used other expressions such as *Not authentic* and *Authentic*.

After identifying the independent constructs among the responses from the three rater groups to the 42 items on the *Student Teachers' Video-taped Instruction* using the Rasch rating scale model, the three groups' responses were analyzed with a series of one-way ANOVAs.

## The Rasch Model

The Rasch measurement model (Rasch, 1960) is an effective means for validating the dimensionality of instruments. The Rasch measurement model allows for an investigation of dimensionality and the ordering of items on an interval scale. The fit of the data to the Rasch measurement model is examined using the item level fit statistics to compare the discrepancy between the observed and model expected responses. Furthermore, the Rasch model reduces complex data matrices to a unidimensional variable, as all systematic variation in the data is explained by one latent variable. The use of Rasch principal component analysis of item residuals provides an effective means of detecting any relevant second construct in the data.

### *Rasch Rating Scale Model*

The Rasch rating scale model (Andrich, 1978; Wright & Masters, 1982) was used to analyze the data and obtain estimates of the degree to which the respondents endorsed the items. The Rasch rating scale model was selected for the data analyses because the items on the *Teaching Skill Questionnaire* used a common rating scale.

The Rasch rating scale model estimates the probability that a respondent will choose a particular response category for an item using the following formula:

$$(P_{nji} / P_{ni}(j - 1)) = B_n - D_i - F_j,$$

where  $P_{nji}$  is the probability of respondent  $n$  selecting category  $j$  for item  $i$ ,

$P_{ni}(j - 1)$  is the probability of respondent  $n$  scoring in category  $j - 1$  of item  $i$ ,

$B_n$  is the person measure of respondent  $n$ ,  $D_i$  is the difficulty of item  $i$ , and

$F_j$  is the difficulty of category step  $j$ .

Rasch analysis places person ability ( $B_n$ ) and item difficulty ( $D_i$ ) on the same measurement scale where the unit of measurement is the logit (logarithm of odds unit). Rasch item and person reliability estimates indicate the reproducibility of the item difficulty estimates and person ability estimates on the scale and is analogous to Cronbach's alpha.

The Rasch logit scale was converted to a CHIPS scale, which sets the mean item difficulty estimate at 50.0; standard errors are generally around 1 CHIP in size. The use of the CHIPS scale makes it easy to estimate of the probability of a participant endorsing a particular item and it eliminates the negative values that are a part of the Rasch logit scale; thus, all measures on the CHIPS scale are positive.

## CHAPTER 6

### RESULTS

#### Research Question 1: Revised COLT Categories

The first research question asked what categories should be included on the observation checklist that student teachers, college teachers, and junior high school teachers (student teachers' supervisors) use when observing student teachers' classroom teaching. The answer to this research question was discussed in Chapter 4.

#### Research Question 2: The *Teaching Skill Questionnaire* Results

Research question 2 asked to what degree student teachers, college teachers and junior high school teachers' ratings on the *Teaching Skill Questionnaire* differed. In order to identify patterns in the responses of the three groups of participants to the 15 items on the *Teaching Skill Questionnaire*, I employed the Rasch rating scale model (Andrich, 1978; Wright & Masters, 1982). Two student teachers failed to complete this questionnaire, so the number of student teacher

respondents was 55. In addition, 19 college teachers and 28 junior high school teachers completed the questionnaire.

I first identified the dimensionality of the responses from the three groups of participants to the 15 items on the *Teaching Skill Questionnaire* using the Rasch rating scale model with Winsteps version 3.68.0 (Linacre, 2009). This was carried by inspecting the pca of item residuals results for the 15 items. A series of analyses indicated that the 15 items were best divided into three groups, with each group measuring a different construct: *Teaching Mainly in Japanese*, *Teaching Using Easy English*, and *Teaching Using Difficult English*.

#### *Question 1-5: Teaching Mainly in Japanese*

The construct underlying questions 1-5 was named *Teaching Mainly in Japanese*. A representative item is “To explain a new structure in Japanese.” The functioning of the original 7-point scale was investigated using Winsteps. Because several thresholds on the scale were extremely close to one another, categories 1, 2, and 3 were collapsed into a single category, as were the 5<sup>th</sup> and 6<sup>th</sup> categories. The new rating scale structure and its functioning are shown in Table 11. The resulting 4-point rating scale met the criteria proposed by Linacre (1999).

Table 11. *Revised Rating Scale Category Structure for Teaching Mainly in Japanese*

Category	Observed (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Easy	51 (12)	1.27	1.29	NONE	(-12.46)
4 Neutral	90 (21)	.81	.80	-6.54	-4.40
5 A little difficult	161 (38)	.86	.78	-1.74	3.71
7 Difficult	120 (28)	1.05	1.06	8.28	(13.63)

As shown in Table 11, the mean Rasch person ability estimates for the college teacher group and the junior high school teacher group were nearly identical at 61.16 and 61.15, respectively. Their means were considerably higher than the mean (51.97) of the student teacher group. The 95% confidence intervals of the college teacher group [56.79, 65.53] and the junior high school teacher group [58.07, 64.23] were much higher and did not overlap with those of the student teacher group [49.83, 54.11], indicating a significant difference between the two sets of scores.

The Rasch item reliability (separation) estimate was 1.00 (15.18), which indicated that the item difficulty estimates were separated widely in relation to their standard errors. The Rasch person reliability (separation) estimate was .45 (.91), indicating that many raters had similar ability estimates and were therefore not well separated on the CHIPS scale.

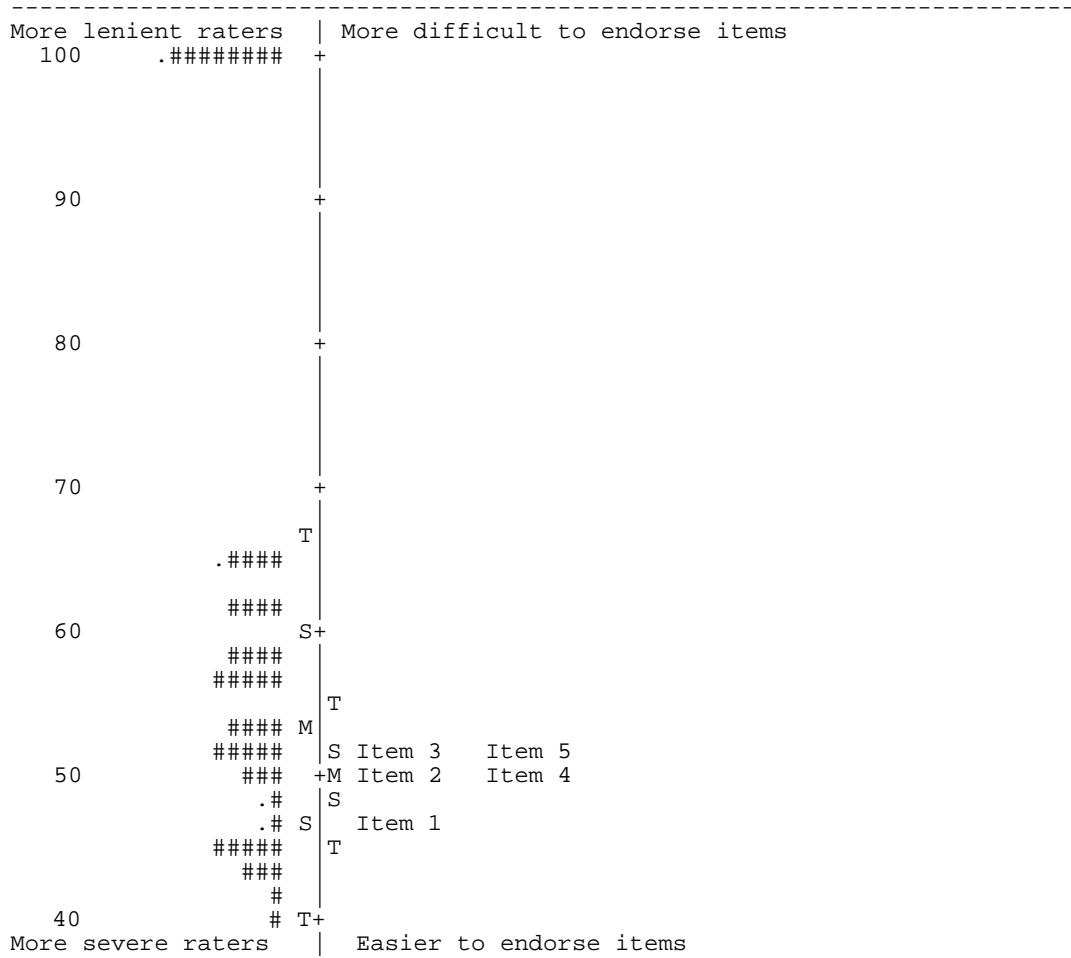
Table 12. *Descriptive Statistics for Teaching Mainly in Japanese*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	51.97	61.16	61.15
<i>SE</i>	1.07	2.09	1.50
95% CI Lower Bound	49.83	56.79	58.07
95% CI Upper Bound	54.11	65.53	64.23
<i>SD</i>	7.91	9.08	7.9
Skewness	.63	-.03	-0.1
<i>SES</i>	0.32	0.53	0.44
Kurtosis	-0.29	-1.66	-1.17
<i>SEK</i>	0.63	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The Wright map for *Teaching Mainly in Japanese* is shown in Figure 1. The mean person ability estimate (*SD*) of 53.20 (6.91) is above the mean item difficulty estimate (*SD*) of 50.00 (3.36), which indicates that the items were somewhat easy to endorse overall. Item 1 (*To greet in English*) was the easiest to endorse (difficulty estimate = 46.2) and items 3 and 5 (*To translate text in the textbook into Japanese; To give background knowledge of the text in the textbook*, respectively) were the most difficult items to endorse (difficulty estimate = 52.4). As shown in Figure 1, the item difficulty estimates clustered together somewhat closely and more difficult to endorse items are needed in order to measure some of the participants more precisely.

The person ability estimates spanned approximately 25 CHIPS, while the item difficulties spanned approximately 6 CHIPS.



Note. Each # is 2 raters. Each . is 1 rater.

Figure 1. Wright map for *Teaching Mainly in Japanese*.

The Rasch model accounted for 46.3% of the variance in the observations.

Unexplained variance (eigenvalues) in the first five contrasts was 18.6% (2.0),

11.3% (1.2), 10.2% (1.1), 5.6% (.6), and 0.1% (.0), respectively. Because of the relatively large amount of variance accounted for by the Rasch model and the relatively small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

Table 13. *Rasch Difficulty and Fit Statistics for Teaching Mainly in Japanese (Items 1-5)*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
1	46.2	.8	1.07	.5	.95	-.2	.70
2	49.4	.8	.71	-2.1	.69	-2.1	.80
3	52.4	.7	1.07	.5	1.11	.8	.78
4	49.5	.8	.84	-1.1	.77	-1.5	.80
5	52.4	.7	1.35	2.2	1.33	2.1	.74

Note. Pt-measure Correlation = Part-measure correlation.

A one-way ANOVA was run to determine whether the student teachers, practicum supervisors, and college teachers' ratings of *Teaching Mainly in Japanese* differed significantly. The independent variable was the three rater groups and the dependent variable was the Rasch person ability estimates. The ANOVA was statistically significant,  $F(2, 99) = 16.01, p = .00$ , partial  $\eta^2 = .25$ . Follow-up tests were conducted to evaluate pairwise differences among the group means. Because the group sizes differed, and the assumption of homogeneity of

variances was met, the Games-Howell procedure was selected. This post hoc test is accurate when sample sizes are unequal (Field, 2005). Statistically significant differences were found in the mean of the ratings between the student teachers and the college teachers ( $p = .001$ ) and between the student teachers and the practicum supervisors ( $p = .001$ ); no statistically significant difference was found between the college teachers and the practicum supervisors ( $p = 1.00$ ). The student teacher group gave significantly lower ratings to *Teaching Mainly in Japanese* than the other two groups. The  $\eta^2$  value, which shows the proportion of the dependent variable that is related to the independent variable, is interpreted in the conventional way in this study; thus, .01, .06, and .14 are interpreted as small, medium, and large effect sizes, respectively (Green & Salkind, 2008). In this case, the effect was large at .25.

#### *Questions 6-10: Teaching Using Easy English*

The construct underlying questions 6-10 was *Teaching Using Easy English*.

A representative item is “To instruct students in Classroom English.”

Table 14. *Revised Rating Scale Category Structure for Teaching Using Easy English*

Category	Observed (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Easy	45 (10%)	1.33	1.20	NONE	(-4.51)
3 A little easy	205 (44%)	.91	.97	-3.40	-1.45
5 A little difficult	147 (31%)	.79	.76	.52	1.71
7 Difficult	70 (15%)	1.09	1.08	2.88	(4.05)

The functioning of the original 7-point Likert scale was investigated using Winsteps. Because a number of points on the scale were found to be extremely close to one another, points 1 and 2, points 3 and 4, and points 5 and 6 were combined. The revised rating scale structure meets all of the criteria proposed by Linacre (1999) (see Table 14).

As shown in Table 15, the mean Rasch person ability estimates of the college teachers and the junior high school teachers are similar at 58.81 and 59.50, respectively, and both means are significantly higher than the mean ability estimate of the student teachers (46.33). The 95% confidence intervals of the college teachers [53.21, 64.40] and the junior high school teachers [55.31, 63.69] show a clear separation from those of the student teachers [43.88, 48.78], indicating significant differences between the student teachers and the other two groups.

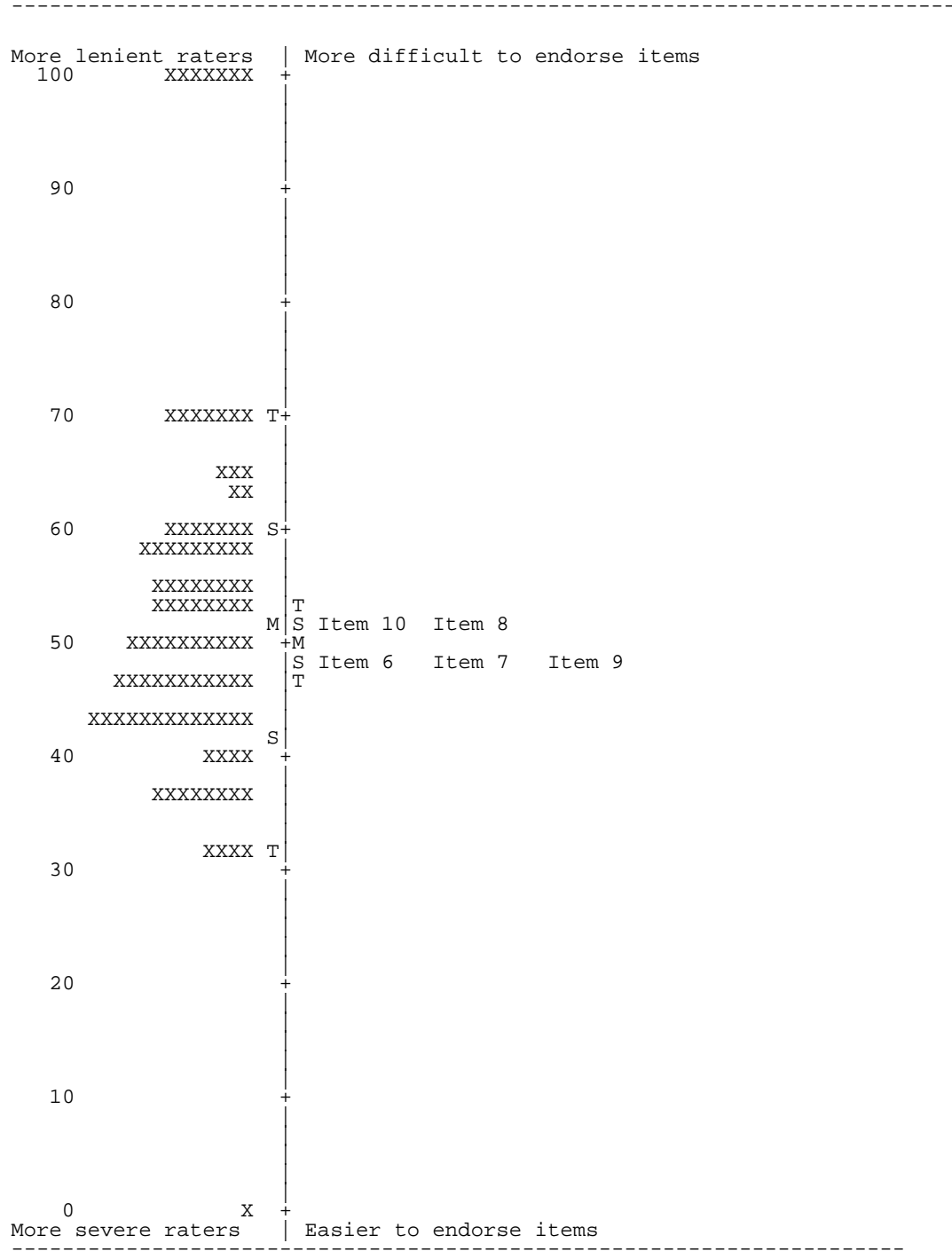
Table 15. *Descriptive Statistics for Teaching Using Easy English*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	46.33	58.81	59.50
<i>SE</i>	1.22	2.66	2.04
95% CI Lower Bound	43.88	53.21	55.31
95% CI Upper Bound	48.78	64.40	63.69
<i>SD</i>	9.07	11.61	10.81
Skewness	-0.14	0.05	-0.16
<i>SES</i>	0.32	0.52	0.44
Kurtosis	0.12	-1.57	-0.16
<i>SEK</i>	0.63	1.01	0.86

*Note.* All statistics are based on Rasch person measure estimates.

The Rasch item reliability (separation) estimate was .62 (1.27), indicating that the item difficulty estimates were not well separated in relation to the standard errors. The Rasch person reliability (separation) estimate was acceptable at .80 (2.02).

The Wright map for *Teaching What Using Easy English* is shown in Figure 2. The mean person ability estimate (SD) of 50.84 (9.98) is almost the same as the mean item difficulty estimate (SD) of 50.00 (1.49), which indicates that the items were generally well centered on the participant ability estimates. Quite a few persons, however, had ability estimates at the top or bottom of the range. Item 6 (*To notice students' serious mistakes in their English and correct them immediately*) was the easiest item to endorse (difficulty estimate = 48.6) and item



Note. Each X is 1 person.  
 Figure 2. Wright map for *Teaching Using Easy English*.

10 (*To make an oral introduction of a new sentence pattern in English*) was the most difficult item to endorse (difficulty estimate = 51.9). As shown in Figure 2, the item difficulty estimates clustered together somewhat closely and both more and less difficult to endorse items are needed in order to measure a number of the participants more precisely.

Table 16. *Rasch Difficulty and Fit for Teaching Using Easy English (Items 6-10)*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
6	48.6	.9	1.68	4.0	1.65	3.6	.71
7	48.8	.9	.73	-2.0	.69	-2.3	.86
8	51.7	.9	.83	-1.2	.82	-1.2	.84
9	49.0	.9	.91	-.6	.91	-.5	.81
10	51.9	.9	.83	-1.2	.81	-1.3	.88

*Note.* Pt-measure Correlation = Part-measure correlation.

The Rasch model accounted for 53.8% of the variance in the observations. Unexplained variance (eigenvalue) in the first five contrasts was 14.7% (1.8), 10.8% (1.4), 8.1% (1.0), 6.2% (.8), and 0.0% (.0), respectively. Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

A one-way ANOVA was conducted to investigate whether there were any statistically significant differences in the means of the ratings of the three groups for *Teaching Using Easy English*. The independent variable was the three groups of raters: the student teachers, the practicum supervisors, and the college teachers. The dependent variable was the ratings of the three groups derived from their responses to items 6-10. The test of homogeneity of variances was significant, so Welch's  $F$  test was used. Welch's  $F$  was significant,  $F(2, 35.73) = 23.29, p = .00$ , partial  $\eta^2 = .31$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the group sizes were different, the Games-Howell procedure was selected, as this post hoc test is accurate when sample sizes are unequal (Field, 2005). There were statistically significant differences in the means between the student teachers and the college teachers ( $p = .001$ ) and between the student teachers and the junior high school teachers ( $p = .001$ ); no statistically significant difference was found between the college teachers and the junior high school teachers ( $p = .98$ ). The student teacher group gave significantly lower ratings to *Teaching Using Easy English*. The student teachers thought it significantly more difficult to teach the points measured with items 6-10.

*Questions 11-15: Teaching Using Difficult English*

The construct underlying questions 11-15 was *Teaching Using Difficult English*. A representative item is “To prepare for the next team-teaching discussing with an ELT in English.”

The functioning of the original 7-point scale was investigated using Winsteps. Because the original rating scale structure met the criteria suggested by Linacre (1999), the rating scale was not altered. The rating scale functioning is shown in Table 17.

Table 17. *Rating Scale Category Structure for Teaching Using Difficult English*

Category	Observed (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Easy	43(9%)	1.27	1.18	NONE	(-3.87)
2 Fairly easy	93 (20%)	.68	.70	-2.60	-2.11
3 A little Easy	111 (24%)	1.01	.92	-1.28	-.88
4 Neutral	106 (23%)	.79	.86	-.43	.15
5 A Little Difficult	51 (11%)	1.06	.99	.88	1.04
6 Fairly Difficult	37 (8%)	.87	.86	1.27	1.99
7 Difficult	30 (6%)	1.69	1.62	2.16	(3.51)

As shown in Table 18, the means of the Rasch ability estimates for the college teachers and the junior high school teachers were 59.93 and 52.30, respectively. Their means were considerably higher than the mean (44.98) of the student teachers. The 95% confidence intervals also showed little or no overlap

between the student teachers [43.46, 46.49] and either the college teachers [51.59, 62.27] or junior high school teachers [49.75, 54.86]. Thus, the two groups of experienced teachers thought that it is easy to teach using difficult English.

The Rasch item reliability estimate (separation) was .88 (2.76), and the Rasch person reliability estimate (separation) was .83 (2.22). Both results indicated good separation of items and persons in relation to their respective errors of measurement.

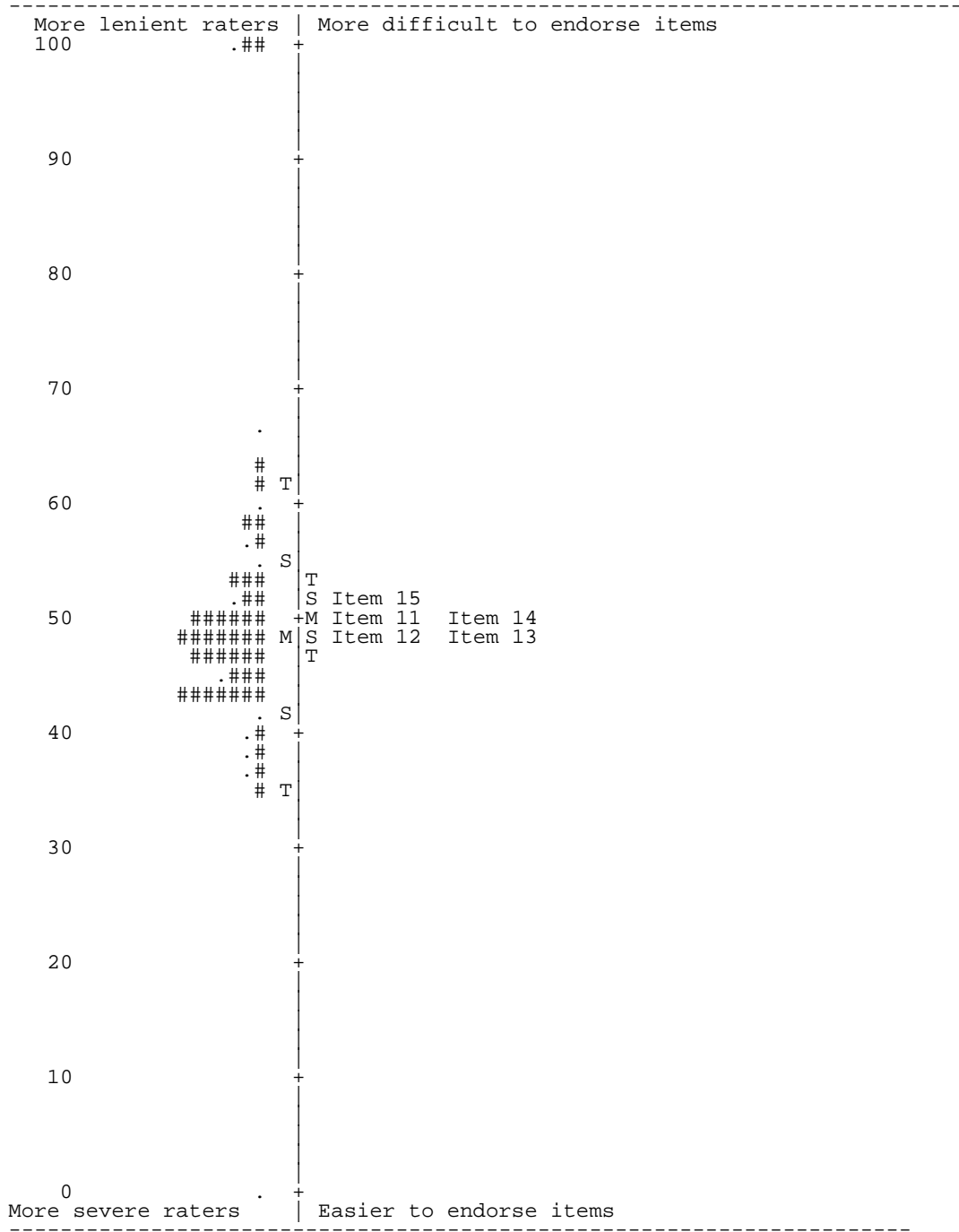
Table 18. *Descriptive Statistics for Teaching Using Difficult English*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	44.98	59.93	52.30
<i>SE</i>	0.75	2.54	1.24
95% CI Lower Bound	43.46	51.59	49.75
95% CI Upper Bound	46.49	62.27	54.86
<i>SD</i>	5.60	11.08	6.58
Skewness	-0.74	-0.23	1.28
<i>SES</i>	0.32	0.52	0.44
Kurtosis	1.93	-1.32	2.26
<i>SEK</i>	0.63	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The Wright map for *Teaching Using Difficult English* is shown in Figure 3.

The mean person ability estimate (SD) of 48.25 (6.46) is below the mean item difficulty estimate (SD) of 50.00 (1.56). This indicates that the items were



Note. Each # is 2 persons. Each . is 1 person.  
 Figure 3. Wright map for *Teaching Using Difficult English*.

somewhat difficult to endorse overall. Item 12 (*To ask questions in English about text and let students try to answer in English*) was the easiest item to endorse (difficulty estimate = 47.8) and items 15 (*To prepare for the next team-teaching discussing with an ELT in English*) were the most difficult items to endorse (difficulty estimate = 52.5). As shown in Figure 3, the item difficulty estimates clustered together somewhat closely and both more and less difficult to endorse items are needed in order to measure the participants more precisely. The lower numbered items were expected to be easier based on the results of the pilot study, but item 11 was a little more difficult than item 12 and item 13.

Table 19. *Rasch Difficulty and Fit Statistics for Teaching Using Difficult English*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
11	50.2	.5	1.32	2.0	1.30	1.9	.82
12	47.8	.5	.97	-.2	.94	-.4	.82
13	49.1	.5	.65	-2.7	.63	-2.8	.88
14	50.4	.5	.85	-1.1	.85	-1.0	.85
15	52.5	.5	1.25	1.6	1.18	1.2	.82

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

The Rasch model accounted for 56.7% of the variance in the observations.

Unexplained variance (eigenvalues) in the first five contrasts was 9.7% (1.5), 9.1% (1.4), 7.2% (1.1), 7.0% (1.1), and 0.0% (.0), respectively. Because of the relatively

large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

Next, a one-way ANOVA was run to investigate differences among the three groups' ratings for *Teaching Using Difficult English*. The homogeneity of variances test was statistically significant, so Welch's  $F$  test was used. The independent variable, raters, was made up of three groups: the student teachers, the practicum supervisors, and the college teachers. The dependent variable was the Rasch person ability estimates derived from items 11-15. Welch's  $F$  was significant,  $F(2, 37.279) = 19.53, p = .00$ , partial  $\eta^2 = .32$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. The Games-Howell procedure was chosen, as this post hoc test is accurate when sample sizes and variances are unequal (Field, 2005). There was a significant difference in the means between the student teachers and the college teachers ( $p = .001$ ) and also between the student teachers and the practicum supervisors ( $p = .001$ ), but no significance between the college teachers and the practicum supervisors ( $p = .25$ ). The  $\eta^2$  value was large at .32. The student teacher group gave significantly lower ratings than the other groups. There was no statistically

significant difference between the college teachers and the junior high school teachers.

### Research Question 3: The Revised COLT

The third research question asked to what degree the student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differed when they used the *Revised COLT*. In order to answer this question, I focused on the ratings the participants awarded to the student teachers' *L2 use* (See Appendix H for the revised COLT) and its related activities in their teaching on the video: The raters awarded of scores 0, 1, or 2 based on how much the student teachers used English as their second language in the class (L2 use) and related activities, such as the purpose of using the L2 (There are two subcategories: *socializing or instruction*, or *content*). For this analysis, I chose one of the subcategories; *socializing or instruction*), the types of questions they asked (There are two subcategories: *pseudo* or *genuine*). For this analysis, I chose one of the subcategories: *pseudo*), the students' use of L2, and sustained speech between student teachers and students (There are two subcategories: *ultra-minimal* or *minimal*). For this analysis I chose one of the categories: *ultra-minimal*).

I selected one teaching situation section from each video segment so that the analysis would be simpler and the participants could observe every student teacher's L2 use. The parts were selected by identifying sections where the student teachers had rich opportunities to use English. Block A showed student teacher 1 teaching *New sentence pattern*, Block B showed student teacher 2 teaching *Introduction of interactive activities*, Block C showed student teacher 3 teaching *Evaluation*, Block D showed student teacher 4 teaching *Model/Chorus reading*, and Block E showed student teacher 5 teaching *Consolidation*.

*L2 use* was not included in Block E; *Purpose of the use of the language (socializing or instruction)* was not included in Block C, and *Sustained speech (ultra-minimal)* was not included in Block D because I had already provided answers for these criteria before distributing the *Revised COLT* sheets to the raters. This was done so that the raters would complete the task more quickly and easily. These data were analyzed with the multi-faceted Rasch model. The ratings that each group awarded the 5 student teachers were then analyzed for group differences with a one-way ANOVA.

### *Student Teacher 1*

As shown in Table 20, the mean of the college teachers, 0.98, was considerably lower than those of the student teacher group (1.84) and the junior high school teacher group (2.90).

Table 20. *Descriptive Statistics for the Three Groups' Ratings for Student Teacher 1*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	1.84	0.98	2.90
<i>SE</i>	0.32	0.60	0.58
95% CI Lower Bound	1.19	-0.28	1.70
95% CI Upper Bound	2.48	2.25	4.10
<i>SD</i>	2.42	2.63	3.09
Skewness	1.47	1.57	1.20
<i>SES</i>	0.32	0.52	0.44
Kurtosis	2.91	4.63	0.25
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch person measure estimates.

The participants observed student 1 on the video and awarded marks of 0, 1, or 2 to categories that they recognized while she was teaching. As shown in Figure 4, *L2 use* was the easiest item and *Purpose* (the use of the language, socializing/instruction or content) was the most difficult item to endorse.

[Measr +student -Rater]		-Item	[Scale]
7	+ ssscjjj  sj	+	+(2)
6	+ +	+	+
5	+ +   ssssjjj	+	+
4	+ +	+	+
3	+ + ssssscccjj	+	+
2	+ +   sssssssssssssssssssssccccjjjjjjjjj	+	+
1	+ +	+	+
* 0 * 1	* sssssssssssssssssssssccccjjjjjjj	* SUSTAINED SPEECH * 1 *	
-1	+ +   sssssccc	+ SL2 USE	+
-2	+ +	+	+
-3	+ + c	+	+
-4	+ +	+	+
		L2 USE	
-5	+ +	+	+(0)
[Measr +student	s=student teacher		
	c=college teacher		
	j=junior high school teacher	-Item	[Scale]

Note. s, c, and j stand for student teacher, college teacher, and junior high school teacher, respectively.

Figure 4. All facets vertical rules for Student Teacher 1.

The assumption of homogeneity of variances was met, and a one-way ANOVA was conducted to investigate differences in the mean of the ratings of the three groups. There was no significant difference among the means,  $F(2, 101) = 3.09, p = .051$ .

### *Student Teacher 2*

The mean of the junior high school teacher group (2.62) was higher than the means of the student teacher group (2.12) and the college teacher group (2.11).

Table 21. *Descriptive Statistics for the Three Groups' Ratings for Student Teacher2*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	2.12	2.11	2.62
<i>SE</i>	0.21	0.44	0.32
95% CI Lower Bound	1.70	1.19	1.96
95% CI Upper Bound	2.53	3.03	3.28
<i>SD</i>	1.57	1.91	1.67
Skewness	0.86	0.50	0.29
<i>SES</i>	0.32	0.52	0.45
Kurtosis	0.42	-0.42	-0.70
<i>SEK</i>	0.62	1.01	0.87

*Note.* All statistics are based on Rasch person measure estimates.

As shown in Figure 5, *SL2 Use* was the easiest item for the raters to observe and notice and *Purpose* the most difficult to endorse. *L2 Use, Question* (Student teachers asking students), and *Sustained speech* were located near the item mean of

0 logits. The raters were located at 0 logits or above indicating that it was difficult for them to endorse this students' performance highly. There was also a ceiling effect with a group of 21 raters located at the top of the logit scale.

A one-way ANOVA was conducted to see if there is any difference in the means of the ratings of the three groups to student 2's performance. There was no significant difference among the means,  $F(2, 101) = .91, p = .41$ .

### *Student Teacher 3*

As shown in Table 22, the means of the three groups of raters differed only slightly. The mean of the student teacher group (4.37) was a little higher than those of the college teacher group (4.17) and the junior high school teacher group (4.18). There was also considerable overlap among the groups' 95% confidence intervals, indicating that the differences in the means are not significant.

*Sustained speech, Question, L2 use, and SL2 use* were endorsed to a similar degree: 49 raters' estimates were located at the top of the scale at 5.0 logits; thus, there was a considerable ceiling effect, which indicated that it was difficult for them to endorse this students' performance highly. A one-way ANOVA was

Measr	+Student	-Rater	-Item	Scale
4	+	ssssssssccc jjjjjjjjj	+	(2)
3	+	sssscccjjj	+	
2	+	ssssssssssssssssssss ccccjjjjjjjjj	PURPOSE	---
1	+	sssssssscccjj	+	
* 0 *	* 2 *	* ssssscccjjj *	SUSTAINED SPEECH	* 1 *
-1	+	ss	QUESTION L2 USE	
-2	+		SL2 USE	(0)
Measr	+student	s=student teacher c=college teacher j=junior high school teacher	-Item	Scale

Note. s, c, and j stand for student teacher, college teacher, and junior high school teacher, respectively.  
 Figure 5. All facets vertical rulers for Student Teacher 2.

Table 22. *Descriptive Statistics for the Three Groups' Ratings for Student Teacher 3*

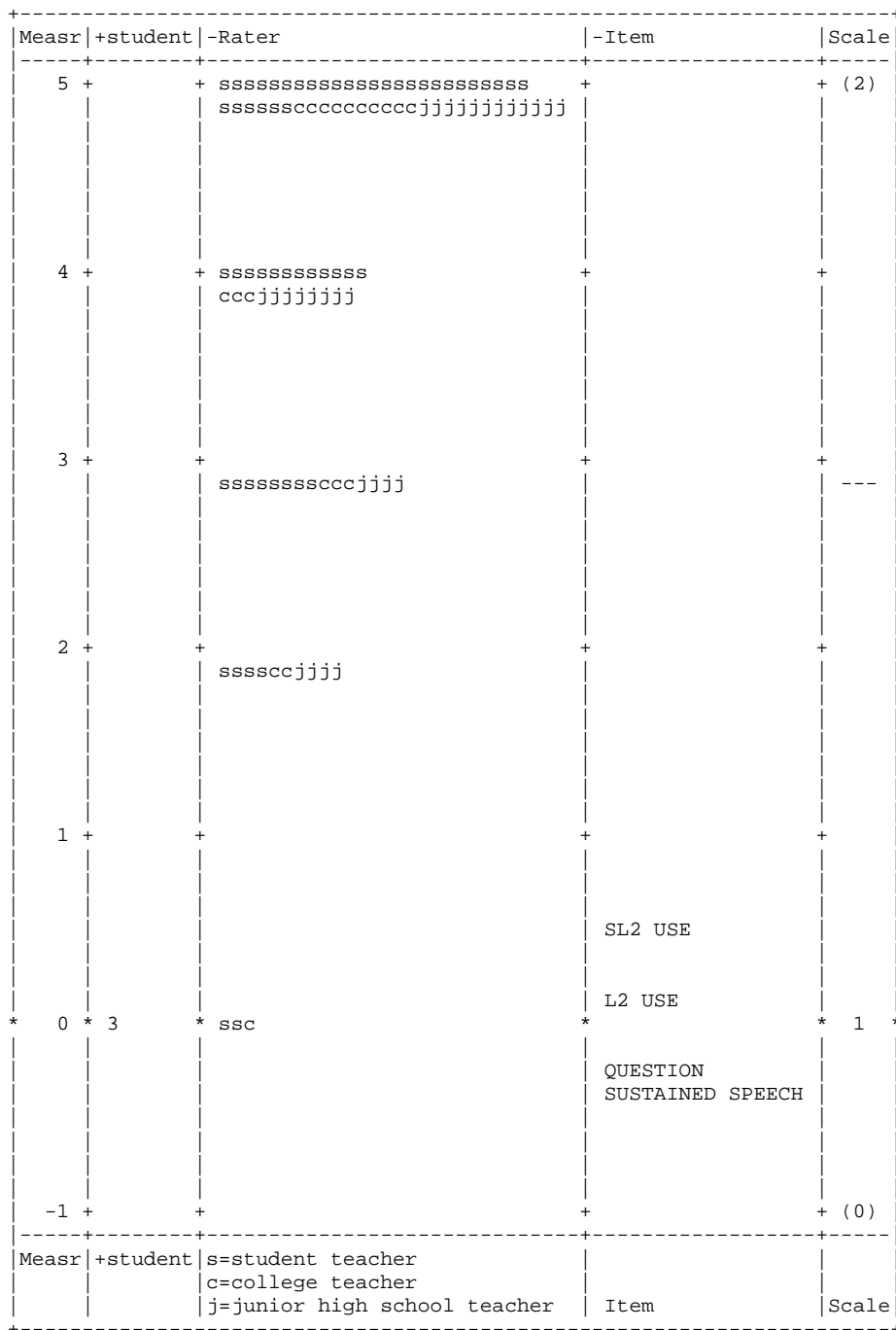
	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	4.37	4.17	4.18
<i>SE</i>	0.19	0.38	0.25
95% CI Lower Bound	3.98	3.37	3.66
95% CI Upper Bound	4.75	4.97	4.70
<i>SD</i>	1.46	1.67	1.34
Skewness	-1.27	-1.11	-0.56
<i>SES</i>	0.32	0.52	0.44
Kurtosis	1.08	0.45	-1.00
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch logits.

conducted to see if there is any difference in the means of the ratings of the three groups to student 3; no significant difference among the means was found,  $F(2, 101) = 0.21, p = .81$ .

#### *Student Teacher 4*

As shown in Table 23, the means of the three groups differed slightly. The mean of the student teacher group (2.50) was a little lower than those of the college teacher group (3.05) and the junior high school teacher group (3.06), but the large degree of overlap in the confidence intervals indicates that the differences are not significant. *SL2 Use* was the easiest item for the participants to endorse, while *Purpose* was the most difficult. *Sustained Speech* was already checked to reduce



Note. s, c, and j stand for student teacher, college teacher, and junior high school teacher, respectively.

Figure 6. Wright map of Student Teacher 3.

Table 23. *Descriptive Statistics for the Three Groups' Ratings for Student Teacher 4*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	2.50	3.05	3.06
<i>SE</i>	0.24	0.44	0.36
95% CI Lower Bound	2.03	2.12	2.32
95% CI Upper Bound	2.98	3.97	3.80
<i>SD</i>	1.79	1.91	1.91
Skewness	0.28	-0.18	0.24
<i>SES</i>	0.32	0.52	0.44
Kurtosis	0.09	0.35	-0.99
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch person measure estimates.

the participants' task (Figure 7.). Most of the raters were located well above 1 logit, which indicated that it was difficult for them to endorse this student's performance highly.

A one-way ANOVA was conducted to determine if there was any difference in the means of the ratings of the three groups for student 4; no significant difference among the means was found,  $F(2, 101) = 1.14, p = .33$ .

#### *Student Teacher 5*

The means of the three groups differ slightly. The mean of the student teacher group was the lowest at 4.21, and that of the junior high school teacher group was

Table 24. *Descriptive Statistics for the Three Groups' Ratings for Student Teacher 5*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	4.21	4.36	4.51
<i>SE</i>	0.16	0.28	0.22
95% CI Lower Bound	3.90	3.78	4.05
95% CI Upper Bound	4.53	4.94	4.97
<i>SD</i>	1.20	1.20	1.19
Skewness	-1.22	-0.65	-1.23
<i>SES</i>	0.32	0.52	0.44
Kurtosis	1.94	-1.48	0.33
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch person measure estimates.

the highest at 4.51. The large degree of overlap in the three groups' 95% confidence intervals indicates that the differences in means are not significant.

*Purpose* and *Sustained speech* were the two easiest criteria for student 5, and *SL2 use* the most difficult. Most of the raters were located near +4 logits and there was a strong ceiling effect, which indicated that it was difficult for them to endorse this students' performance highly. *L2 Use* was already checked to reduce the raters' talk (Figure 8).

A one-way ANOVA was conducted to see if there is any difference in the means of the ratings of the three groups to the student 5. No significant difference among the means was found,  $F(2, 101) = 0.60, p = .55$ .

Measr	+student	-Rater	-Item	Scale
5	+	+	+	+
		sssssc		(2)
		jjjj		
		ssssccjjjj		
4	+	+	+	+
3	+	+	+	+
		ssssssssssssss	PURPOSE	---
		ssccccjjjjjj		
2	+	+	+	+
		ssssssssssssss		
		ssccccjjjjjjjj		
1	+	+	+	+
* 0	* 4	*	*	* 1 *
		sssjj	QUESTION	
			L2 USE	
-1	+	+	+	+
		ssc		
-2	+	+	+	+
			SL2 USE	(0)
Measr	+student	s=student teacher c=college teacher j=junior high school teacher	Item	Scale

Note. s, c, and j stand for student teacher, college teacher, and junior high school teacher, respectively.

Figure 7. All facets vertical rulers for Student Teacher 4.

Measr	+student	-Rater	-Item	Scale
4	+	+ ssssssssssssssssssssscc+		+
		ccccccccjjjjjjjjjjjjjjjj		(2)
		ssssssssssssss		
		ssssssccjjjj		
3	+	+	+	+
		ssssccc		---
		ccjjj		
2	+	+	+	+
		ssj	SL2 USE	
1	+	+	+	+
0	* 5	* s	* QUESTION	* 1 *
			PURPOSE SUSTAINED SPEECH	
-1	+	+	+	+
				(0)
Measr	+student	s=student teacher	Item	Scale
		c=college teacher		
		j=junior high school teacher		

Note. s, c, and j stand for student teacher, college teacher, and junior high school teacher, respectively.

Figure 8. All facets vertical rules for Student Teacher 5.

## The Student Teachers' Videotaped Instruction

The fourth research question asked to what degree the student teachers, college teachers, and junior high school teachers' ratings of student teachers' teaching performances differed when they used the *Student Teachers' Videotaped Instruction*. The Rasch rating scale model was used for the analyses.

I divided the 42 questions of the *Student Teachers' Videotaped Instruction* into three categories. The first category was made up of 22 questions that asked about the usefulness of teaching skills and theories. These were measured on a 5-point Likert scale. The second category was made up of 16 questions concerning the effectiveness of the student teachers' teaching, most of which were paired with some of the 22 questions. For example, one of the 22 questions concerning usefulness was: *Is it useful if a teacher reviews the class in English trying not to use Japanese?* The paired question concerning effectiveness was: *The student teacher on the screen tries to review the class in English without using Japanese. Is she effective when her English and teaching technique are taken into consideration?* The third category was made up of four questions based on differently worded items. For example, one of them provided the raters with five

choices: 1 = *Very authentic*, 2 = *Somewhat authentic*, 3 = *Neutral*, 4 = *Not very authentic*, 5 = *Not at all authentic*.

### *Very Useful Or Not Very Useful Likert Items*

I first identified independent constructs among the responses from the three rater groups to the items designed to measure the usefulness of teaching theories related with English education in Japanese junior high schools on the *Student Teachers' Videotaped Instruction* using the Rasch rating scale model. Four constructs were identified: *Listening and Japanese Use*, *Prereading*, *Using English*, and *Memorization*.

#### *Construct 1: Listening and Japanese Use*

The construct underlying question 5, 7, 13, 30, and 39 was named *Listening and Japanese Use*. An example item, question 7 was, "Is it useful to explain English grammar in Japanese?"

When category functioning was inspected, it was found that category 5 was used only 17 times (3%), so categories 4 and 5 were combined. The resulting rating scale met the criteria outlined by Linacre (1999) (Table 25).

Table 25. *Rating Scale Functioning for Listening and Japanese Use*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not useful	44 (9)	1.22	1.10	NONE	(-17.51)
2 Not very useful	121 (24)	.87	.84	-12.42	-4.45
3 Neutral	119 (23)	.98	1.30	4.03	6.28
4 Useful	220 (44)	.97	.99	8.38	(14.37)

The means of the student teacher group, the college teacher group, and the junior high school teacher group are similar at 57.44, 57.33, and 54.56, respectively.

The 95% confidence intervals overlap considerably, indicating that the differences are not statistically significant.

Table 26. *Descriptive Statistics for the Listening and Japanese Use Items*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	57.44	57.33	54.56
<i>SE</i>	1.18	1.45	1.07
95% CI Lower Bound	55.08	54.28	52.36
95% CI Upper Bound	59.80	60.37	56.77
<i>SD</i>	8.90	6.32	5.68
Skewness	-1.18	0.31	0.21
<i>SES</i>	0.32	0.52	0.44
Kurtosis	3.54	-1.49	0.46
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The Rasch item reliability (separation) was very good at .99 (10.35), and the Rasch person reliability (separation) was .49 (.99); thus, the person ability estimates were not well separated in relation to their standard errors.

Table 27. *Rasch Descriptive Statistics for the Listening and Japanese Use Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
Question 5	66.9	.90	1.23	1.50	1.22	1.40	.54
Question 13	45.2	.80	1.07	.50	.97	.00	.67
Question 30	45.0	.80	.89	-.70	1.02	.20	.67
Question 39	48.3	.70	.81	-1.50	1.00	.60	.72
Question 7	44.7	.80	.88	-.80	.94	-.10	.67

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

The person mean (*SD*) of 56.89 (6.90) was considerably higher than the item mean (*SD*) of 50.00 (8.53), indicating that the items were somewhat difficult to endorse. Question 7 (“Is it useful to explain English grammar in Japanese?”) was the easiest item to endorse (difficulty estimate = 44.7) and question 5 (“In these two scenes, what the students did was just listening except some occasions. Is this useful for review?”) was the most difficult (difficulty estimate = 66.9).

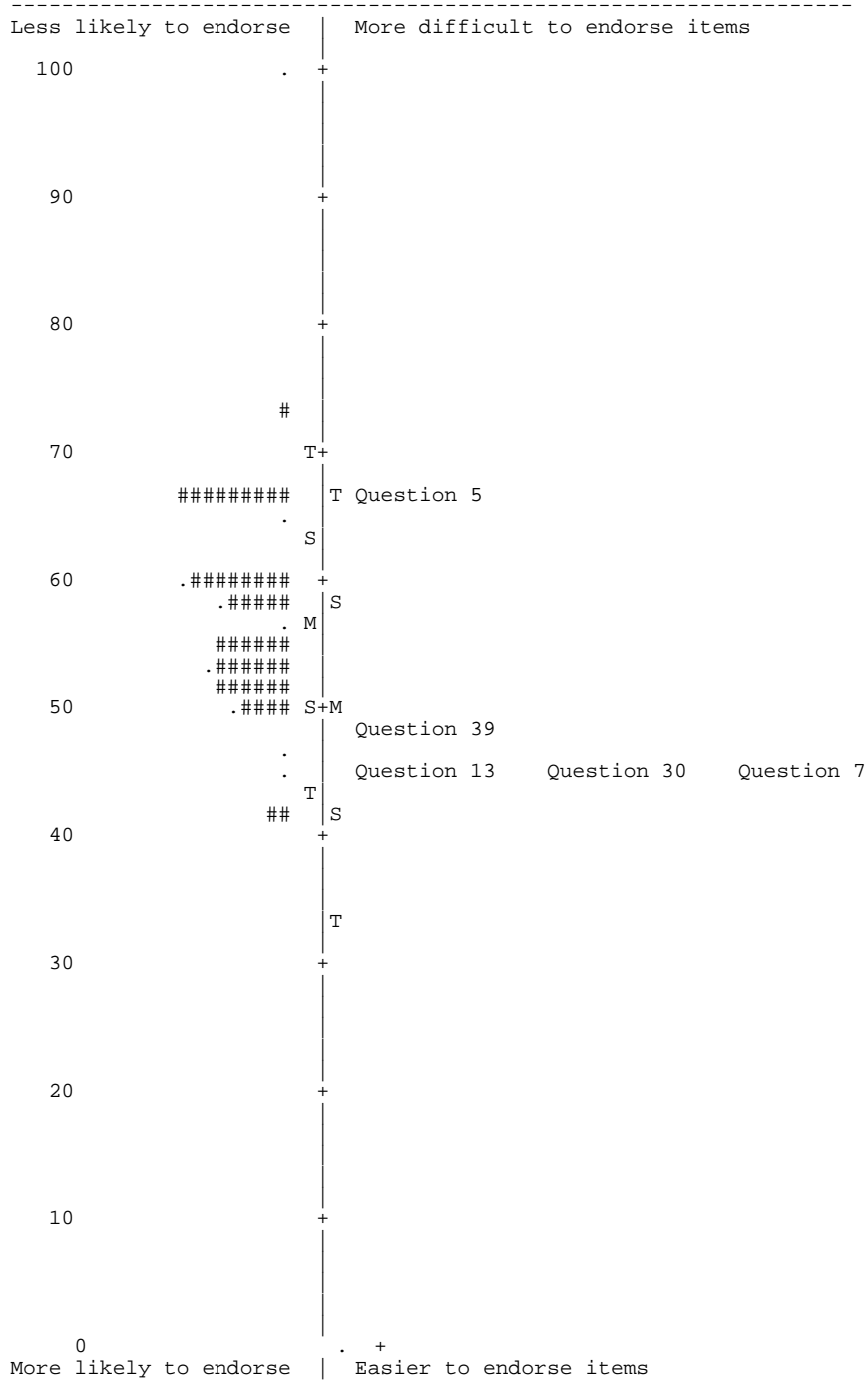
The Rasch model accounted for 69.1% of the variance in the observations (eigenvalue = 11.2). Unexplained variance in the first contrast was 9.2% (eigenvalue = 1.5). The eigenvalues for the remaining contrasts were < 1.5.

Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the contrasts, it was concluded that the construct was fundamentally unidimensional.

### *Construct 2: English Use*

The construct underlying questions 1, 3, 11, 19, 33, and 35 was named *English Use*. For example, question 3 asked “Is it useful if a teacher introduces a new sentence pattern in English trying to lead students to find out a new rule?” The functioning of the original 5-point scale was investigated using Winsteps. Because the thresholds for categories 1 and 2, and categories 4 and 5 were extremely close to one another, each pair was collapsed into a single category. The new rating scale structure and its functioning are shown in Table 28. The resulting 4-point rating scale met the criteria proposed by Linacre (1999).

The means of the student teachers, college teachers and junior high school teachers were 51.81, 49.50, and 52.95, respectively. The 95% confidence intervals overlap considerably, indicating that the differences are not statistically significant.



Note. Each '#' is 2 raters.  
 Figure 9. Wright map for *Listening and Japanese Use*.

Table 28. *Rating Scale Functioning for English Use*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not useful	106 (18)	1.17	1.22	NONE	(-14.37)
3 Useful	336 (56)	.90	1.02	-9.32	.00
4 Very useful	154 (25)	.89	.87	9.32	(14.37)

Table 29. *Descriptive Statistics for English Use*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	51.81	49.50	52.95
<i>SE</i>	1.38	1.40	1.31
95% CI Lower Bound	49.05	46.55	50.24
95% CI Upper Bound	54.57	52.45	55.66
<i>SD</i>	10.40	6.12	6.98
Skewness	-.06	0.13	-0.69
<i>SES</i>	0.316	0.52	0.44
Kurtosis	.02	0.19	0.51
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The Rasch item reliability estimate (separation) was .96 (4.65), and the Rasch person reliability estimate (separation) was .67 (1.41); thus, the person ability estimates were not well separated in relation to their standard errors.

The person mean (*SD*) was 52.26 (7.39), and the item mean (*SD*) was 50.00 (4.15), indicating that these items were difficult to endorse. Question 19 (“Is it useful if a language game is interesting to students?”) was the easiest item to endorse

Table 30. *Rasch Descriptive Statistics for the English Use Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-Measure Correlation
Question1	56.0	1.0	1.27	1.9	1.38	2.2	.64
Question3	55.1	1.0	.75	-1.9	.74	-1.7	.74
Question11	50.2	1.0	.79	-1.6	.80	-1.4	.63
Question19	42.0	1.0	1.16	1.2	1.10	.5	.69
Question33	50.1	1.0	.93	-.4	.95	-.3	.63
Question35	46.6	1.0	1.16	1.2	1.10	.5	.69

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

(difficulty estimate = 42.0) and question 1 (“Is it useful if a teacher reviews the class in English trying not to use Japanese?”) was the most difficult (difficult estimate = 56.0).

The Rasch model accounted for 49.6% of the variance in the observations (eigenvalue = 5.9), and the unexplained variance in the first contrast was 13.9% (eigenvalue = 1.6). The eigenvalues for the remaining contrasts were equal to or less than 1.4. Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional

### *Construct 3: Prereading*

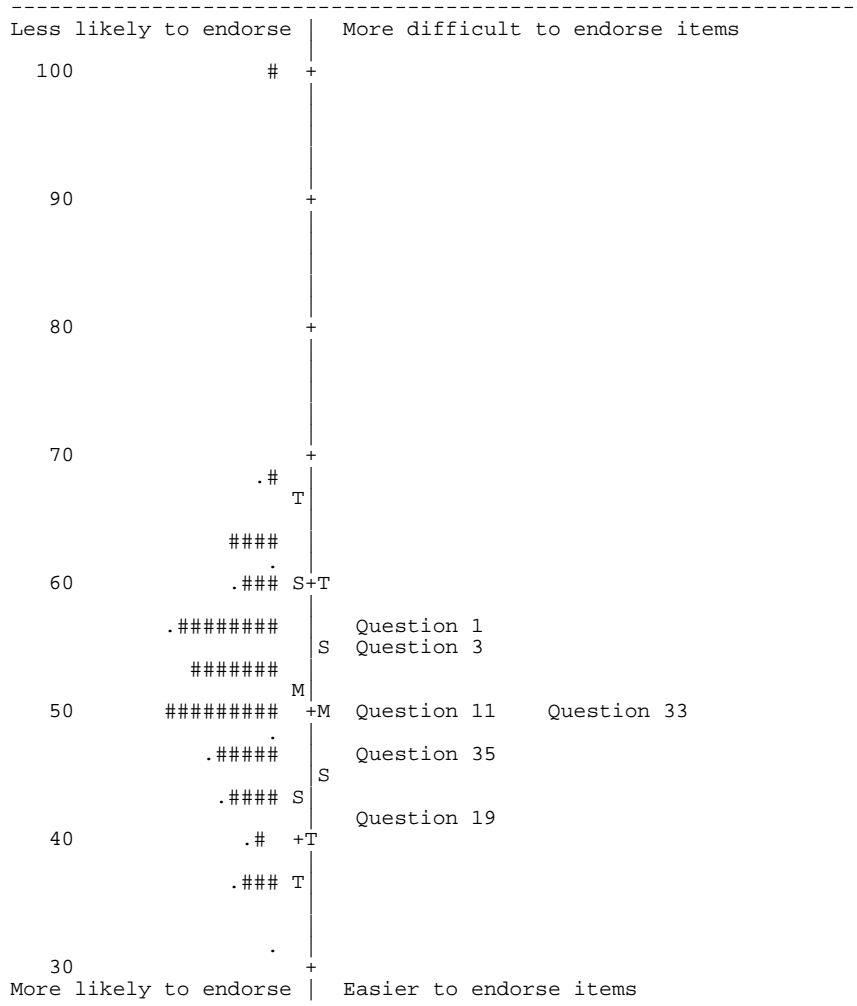
The construct underlying questions 6, 24, and 36 was *Prereading*. Question 6 is “In these two scenes, the students just listened except on some occasions. Is this enough for introducing the new sentence pattern?”

The functioning of the original 5-point scale was investigated using Winsteps.

Category 1 was selected only 19 times (6%), so categories 1 and 2 were combined (Table 31). The resulting 4-point rating scale met the criteria proposed by Linacre (1999).

The rating scale structure and its functioning are shown in Table 31. As shown in Table 33, all three items had high part-measure correlations, they fit the Rasch model well, and the item difficulty estimates were well separated.

The means of the student teachers, college teachers, and junior high school teachers were very similar at 50.59, 49.10, and 48.94, respectively (See Table 32). The standard deviation of the junior high school teachers was somewhat smaller than that of the student teachers and college teachers, indicating less variability in the responses of the junior high school teachers.



Note. Each '#' is 2 raters.  
Figure 10. Wright map for *English Use*.

Table 31. *Rating Scale Functioning for Prereading*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not useful	70 (23)	.98	1.05	NONE	(-14.15)
2 Neutral	45 (15)	.83	.93	-7.77	-6.72
3 Useful	140 (47)	.85	.94	-5.56	4.19
4 Very useful	46 (15)	1.14	1.12	17.85	(22.91)

Note. The neutral category was originally the mid-point (3) of the 5-point Likert scale.

Table 32. *Rasch Descriptive Statistics for the Prereading Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-Measure Correlation
Question 36	45.60	.80	.93	-.40	1.10	.60	.63
Question 24	42.60	.90	1.03	.20	.96	-.10	.66
Question 6	61.80	.80	.86	-.90	.93	-.20	.73

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

The Rasch item reliability estimate (separation) was very good at .99 (10.05), and the Rasch person reliability estimate (separation) was .34 (.72); thus, the person ability estimates were not well separated in relation to their standard errors. The person mean (*SD*) was 50.02 (7.22), and the item mean (*SD*) was 50.00 (7.96), indicating that the items were well matched with the participants. Question 24

Table 33. *Descriptive Statistics for Prereading*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	50.59	49.10	48.94
<i>SE</i>	1.16	1.96	1.24
95% CI Lower Bound	48.25	44.98	46.39
95% CI Upper Bound	52.92	53.22	51.50
<i>SD</i>	8.80	8.55	6.57
Skewness	-0.18	2.60	0.68
<i>SES</i>	0.32	0.52	0.44
Kurtosis	0.26	8.91	2.07
<i>SEK</i>	0.62	1.01	0.86

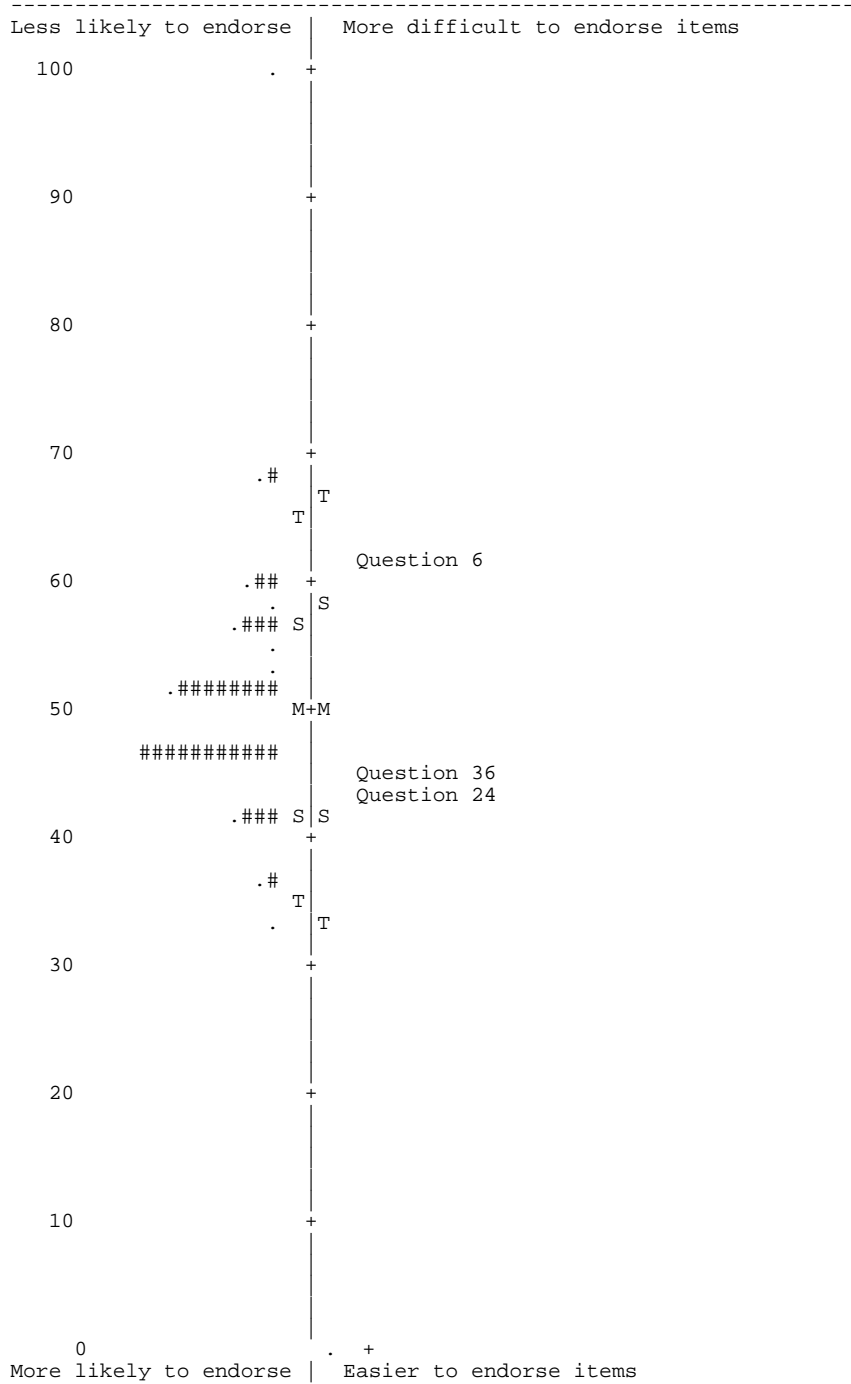
*Note.* All statistics are based on Rasch CHIPS.

“Is it useful to introduce new words with such as flash cards checking the pronunciation, the meanings, and such useful before going into a text?”) was the easiest item (difficulty measure = 42.60) and question 6 (“In these two scenes, what the students did was just listening except some occasions. Is this useful for introduction of the new sentence pattern?”) was the most difficult item (difficulty measure = 61.80).

The Rasch model accounted for 72.3% of the variance in the observations (eigenvalue = 7.8). Unexplained variance in the first contrast was 13.9%, and eigenvalue was 1.5. The eigenvalues for the remaining contrasts were equal to or less than 1.4. Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

#### *Construct 4: Memorization*

The construct underlying questions 9 and 37 was named *Memorization*. Question 9 asked, “In order to learn a new sentence pattern, is this useful for students to repeat the new pattern after the teacher?” and question 37 asked, “Is it useful to let students memorize the text as part of post reading activities?”



Note. Each '#' is 3 raters.  
 Figure 11. Wright map for Prereading.

Category 1 was not chosen by any of the respondents. Category 2 was used only 8 times (4%), so categories 1 and 2 were combined. Category 5 was used only 11 times (6%), so categories 4 and 5 were combined. The resulting categories met the criteria proposed by Linacre (1999) (see Table 34).

Table 34. *Rating Scale Functioning for Memorization*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not useful	33 (19)	.92	.94	NONE	(-25.11)
2 Useful	128 (74)	1.01	1.01	-20.12	.00
3 Very useful	11 (6)	1.12	1.06	20.12	(25.11)

The means of the student teachers (44.57), and college teachers (44.01) were lower than that of the junior high school teachers (49.84). However, there was considerable overlap among the 95% confidence intervals for the three groups, indicating that the differences were not statistically significant.

The Rasch item reliability estimate (separation) was very good at .96 (4.65); however, the Rasch person reliability estimate (separation) was .00 (.00); thus, the person ability estimates were not well separated in relation to their standard errors. This is due in large part to the fact that this construct was measured with only two items.

Table 35. *Descriptive Statistics for Memorization*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	44.57	44.01	49.84
<i>SE</i>	2.23	4.47	3.23
95% CI Lower Bound	40.11	34.69	43.20
95% CI Upper Bound	49.03	53.49	56.47
<i>SD</i>	16.80	19.51	17.11
Skewness	0.49	0.51	0.49
<i>SES</i>	0.32	0.52	0.44
Kurtosis	-0.36	-0.29	0.6
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

Table 36. *Rasch Descriptive Statistics for the Memorization Items*

Item	Measure	<i>SE</i>	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-Measure Correlation
Question 9	42.30	1.60	1.00	.10	.99	.10	.80
Question 37	57.70	1.60	1.00	.10	1.0	.10	.88

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

The person mean (*SD*) was 44.86 (12.33), and the item mean (*SD*) was 50.00 (7.72), indicating that these items were somewhat easy to endorse.

The Rasch model accounted for 61.8% of the variance in the observations (eigenvalue = 3.2). Unexplained variance in the first contrast was 0.0%, and eigenvalue was 0.0. The eigenvalues for the remaining contrasts were 0.0. Because of the relatively large amount of variance accounted for by the Rasch model and the

small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

Next, the responses of the three groups to each construct were analyzed with a series of one-way ANOVAs in order to answer research question 4: To what degree do student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differ when they use the *Student Teachers' Videotaped Instruction*?

#### *Listening and Japanese Use*

A one-way ANOVA was conducted to investigate the differences in the means of the ratings of the three groups to the *Listening and Japanese Use* construct. The independent variable was the three groups and the dependent variable was the ratings of the three groups. The result was not statistically significant,  $F(2, 101) = 1.39, p = .25$ .

#### *English Use*

The test of homogeneity of variances was significant for the *English Use* construct, so Welch's  $F$  test was conducted to determine if there were any

significant differences in the mean ratings of the three groups to the *English Use* construct. The result was not significant,  $F(2, 53.95) = 1.62, p = .21$ .

### *Prereading*

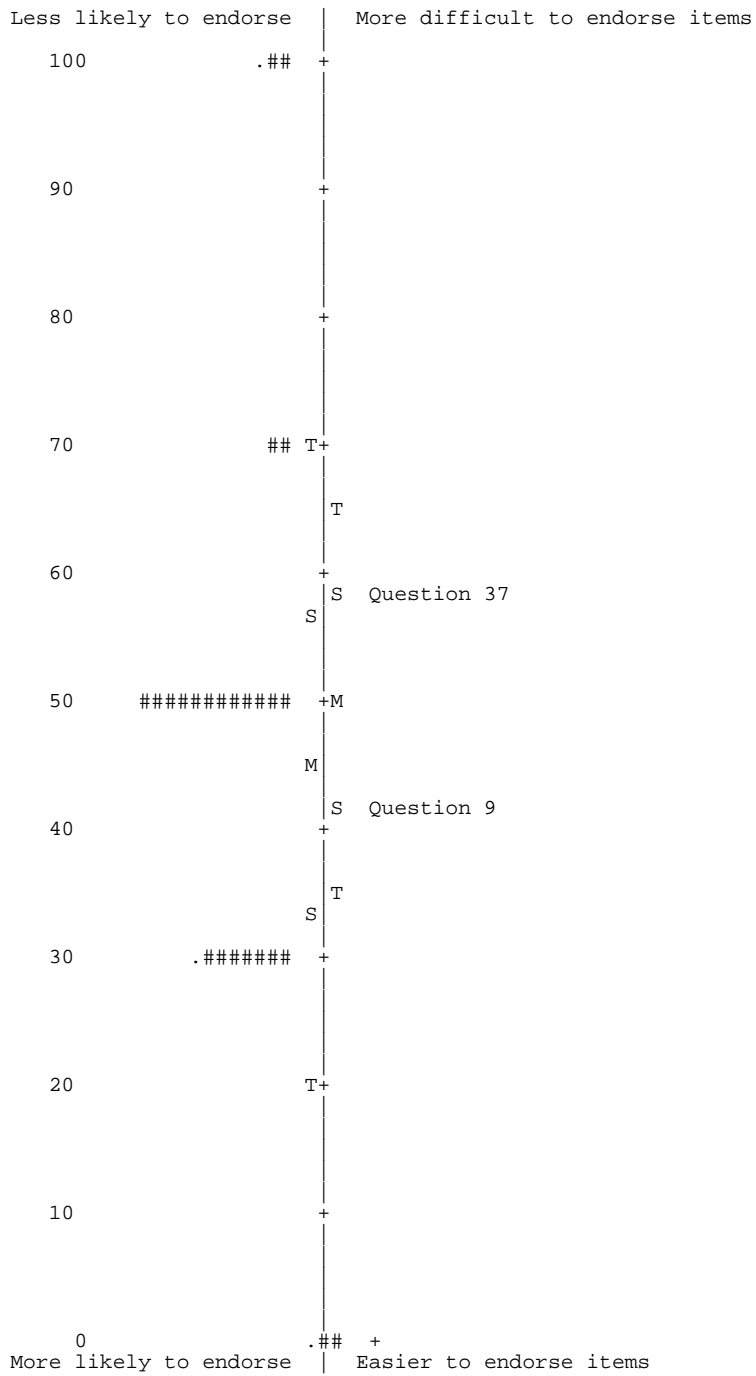
A one-way ANOVA was run to determine if there were any significant differences in the mean ratings of the three groups to the *Prereading* construct. The results were not significant,  $F(2, 101) = .48, p = .62$ .

### *Memorization*

A one-way ANOVA was conducted in order to determine if there were any significant differences in the means of the ratings of the three groups to the *Memorization* construct. There were no significant differences,  $F(2,101) = .99, p = .38$ .

### *The Effectiveness of the Student Teachers' Instruction*

I first identified independent constructs among the responses from the three rater groups to the items in this section of the *Student Teachers' Video Instruction*



Note. Each '#' is 4 raters.  
 Figure 12. Wright map for Memorization.

using the Rasch rating scale model. Three constructs were identified: *Explanation and Understanding*, *Practice and Habit-formation*, and *Communicative Practice*.

*Construct 1: Explanation and Understanding*

The construct underlying question 2, 4, 12, 16, 27, 29, and 42 was named *Explanation and Understanding*. One example question asked, “Is the student teacher’s instruction before and after the game effective?” In this video clip the student teacher reminds her students of eye-contact, intonation, and other issues involved in face to face communication and explains how to play the game. After the language game, she asks the students to assess their eye contact, intonation, and the other issues she discussed.

When the rating scale functioning was checked, it was found that Category 1 was selected only 18 times (3%), so categories 1 and 2 were combined. As shown in Table 37, the resulting rating scale met the criteria described by Linacre (1999).

Table 37. *Rating Scale Functioning for Explanation and Understanding*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not effective	172 (24)	.85	.89	NONE	(-13.87)
3 Neutral	196 (28)	1.01	1.00	-7.77	-6.02
4 Effective	300 (42)	1.03	1.12	-4.09	4.19
5 Very effective	44 (6)	1.20	1.10	11.86	(16.96)

The means of the student teachers, college teachers and junior high school teachers were 44.28, 44.89, and 44.78, respectively. The 95% confidence intervals indicated that the student teacher group [46.70, 49.87] differed meaningfully from the college teachers group [42.67, 47.11] and junior high school teacher group [43.26, 46.30].

Table 38. *Descriptive Statistics for Explanation and Understanding*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	44.28	44.89	44.78
<i>SE</i>	0.79	1.06	0.74
95% CI Lower Bound	46.70	42.67	43.26
95% CI Upper Bound	49.87	47.11	46.30
<i>SD</i>	5.98	4.06	3.93
Skewness	-0.19	0.10	-0.30
<i>SES</i>	0.32	0.52	0.44
Kurtosis	3.47	-1.08	-0.01
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The Rasch item reliability estimate (separation) was very good at .97 (5.94), indicating that the item difficulty estimates were well separated in relation to the standard errors. The Rasch person reliability estimate (separation) was low at .64 (1.33).

The person mean (*SD*) of 46.93 (5.10) was somewhat below the item mean (*SD*) of 50.00 (4.42), indicating that the items were somewhat easy for the raters to

Table 39. *Rasch Descriptive Statistics for the Understanding and Explanation*

*Items*

Item	Measure	<i>SE</i>	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
Question 2	57.4	.7	1.20	1.4	1.24	1.3	.57
Question 4	55.7	.7	.81	-1.5	.81	-1.2	.67
Question 12	48.7	.7	.76	-1.9	.81	-1.4	.60
Question 16	48.3	.7	1.04	.3	1.03	.3	.54
Question 27	49.4	.7	.80	-1.7	.82	-1.4	.61
Question 29	45.2	.7	1.31	2.0	1.35	2.1	.43
Question 42	45.3	.7	1.05	.4	1.07	.5	.63

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

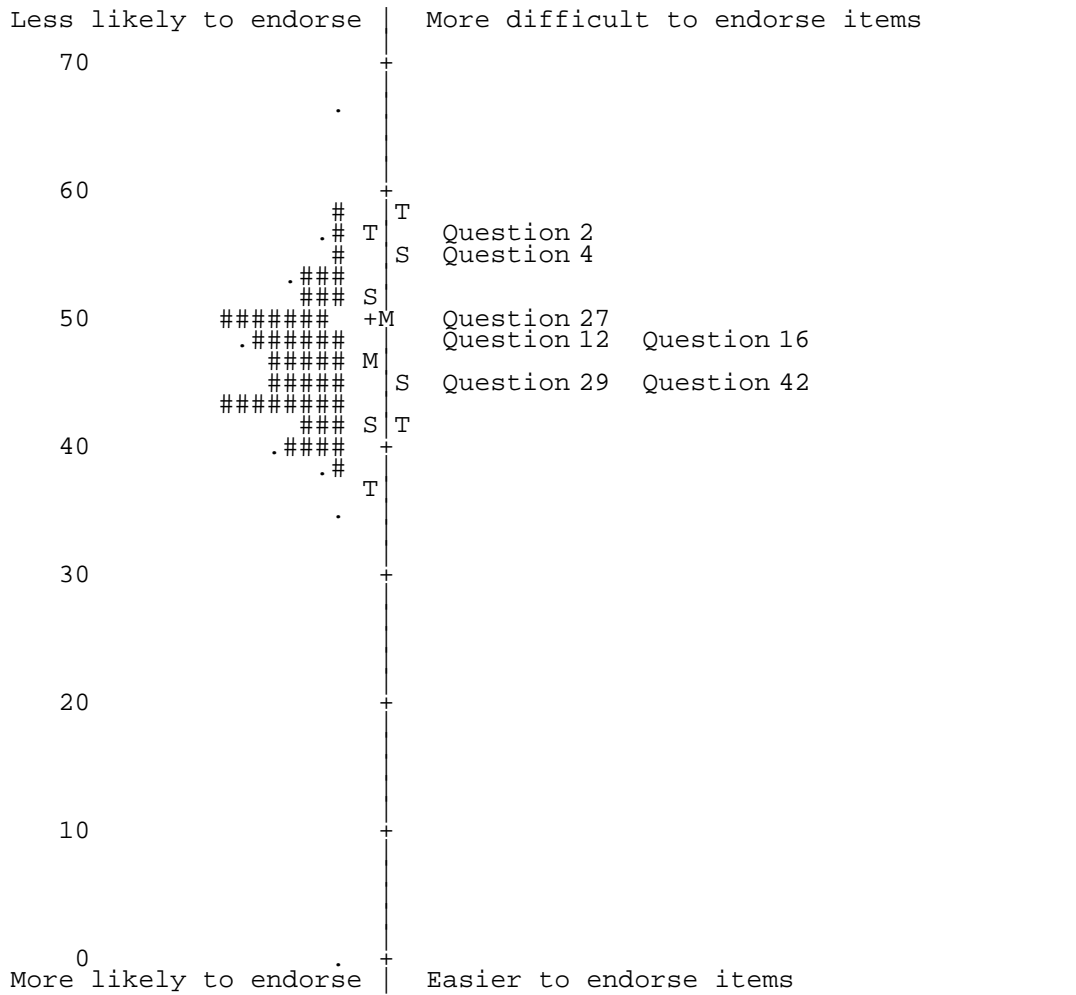
endorse. Question 29 (*The student teacher let the students underline sentences which have the new sentence pattern so that they will be aware of the new sentence pattern in the text. Is she effective?*) was the easiest item to endorse (difficulty measure = 45.2) and question 2 (*The student teacher on the screen tries to review the class in English trying not to use Japanese. Is she effective when her English?*) was the most difficult (difficulty measure = 57.4). What made Question 2 (57.4), 4 (55.7), and 27 (49.4) different from the other questions 12, 16, 42, and 29 (48.7, 48.3, 45.3, and 45.2, respectively) was that in the activities of Question 2, 4, and 27, the student teachers used only English

The Rasch model accounted for 50.1% of the variance in the observations (eigenvalue = 7.0). Unexplained variance in the first contrast was 14.2%, and the eigenvalue was 2.0. The eigenvalues for the remaining contrasts were equal to or less than 1.4. Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

A one-way ANOVA was conducted to determine if there were any significant differences in the ratings of the three groups. The independent variable, the raters, was made up of three groups: the student teachers, the college teachers, and the junior high school teachers. The dependent variable was the ratings for the *Explanation and Understanding* construct. The ANOVA was significant,  $F(2, 101) = 5.56, p = .005$ , partial  $\eta^2 = .10$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the group sizes were different but the test of homogeneity of variances was not significant, the Games-Howell procedure was used. There was a statistically significant difference in the means between the student teachers and the college teachers ( $p = .036$ ), and between the student teachers and the junior high teachers ( $p = .005$ ). No significant difference was found between the college

teachers and junior high school teachers ( $p = .996$ ). The  $\eta^2$  value was large at .10.



Note. Each '#' is 2 persons.

Figure 13. Wright map for *Explanation and Understanding*.

### Construct 2: Practice and Habit Formation

The construct underlying questions 8, 10, 14, 25, and 40 was named *Practice and Habit Formation*. A sample item is: “The student teacher let the students translate the text into Japanese. Is she effective?”

The category functioning was checked and it was found that category 1 was used only 5 times (1%), so categories 1 and 2 were combined. As shown in Table 40, the resulting category structure met the criteria proposed by Linacre (1999).

Table 40. *Rating Scale Functioning for Practice and Habit Formation*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not effective	100 (20)	1.03	1.02	NONE	(-15.50)
2 Neutral	157 (31)	1.10	1.20	-9.64	-7.09
3 Effective	227 (44)	.88	.88	-4.45	5.00
4 Very effective	24 (5)	1.00	.88	14.09	(19.15)

The mean of the junior high school teacher group (41.93) was lower than those of the college teacher group (47.95) and the student teacher group (48.61).

Table 41. *Descriptive Statistics for Practice and Habit Formation*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	48.61	47.95	41.93
<i>SE</i>	0.77	2.04	1.03
95% CI Lower Bound	47.07	43.31	39.82
95% CI Upper Bound	50.15	51.87	44.03
<i>SD</i>	5.81	8.88	5.43
Skewness	0.03	1.18	-0.55
<i>SES</i>	0.32	0.52	0.44
Kurtosis	-0.25	2.20	0.80
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch CHIPS.

The 95% confidence intervals indicated that the junior high school teachers' ratings [39.82, 44.03] were considerably lower than those of the student teachers [47.07, 50.15] and college teachers [43.31, 51.87].

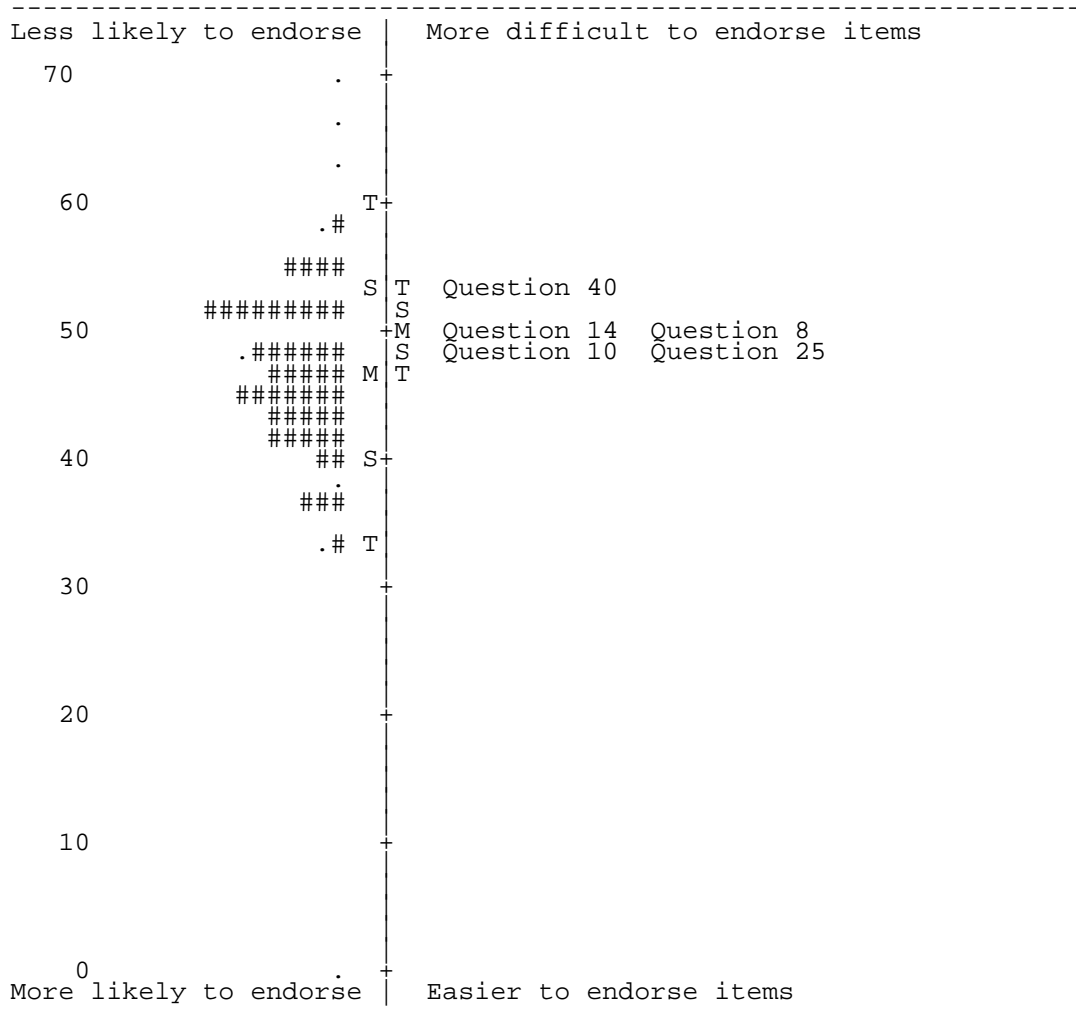
The Rasch item reliability estimate (separation) was somewhat low at .75 (1.75), and the Rasch person reliability estimate (separation) was low at .66 (1.39).

Table 42. *Rasch Descriptive Statistics for the Practice and Habit Formation Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
Question 8	49.9	.7	.64	-3.1	.59	-3.2	.78
Question 10	47.9	.7	.98	-.1	.96	-.2	.73
Question 14	50.5	.7	1.32	2.2	1.35	2.2	.54
Question 25	49.2	.7	1.09	.7	1.03	.2	.64
Question 40	52.5	.7	.93	-.5	1.09	.6	.64

*Note.* Pt-measure Correlation = Part-measure correlation.

The person mean (*SD*) of 46.82 (6.68) was somewhat lower than the item mean (*SD*) of 50.00 (1.52), indicating that the items were somewhat easy for the raters to endorse. Question 10 (“In order for learners to learn a new sentence pattern, the student teacher on the screen let her students repeat after her. Is she effective?”) was the easiest item to endorse and question 40 (“The student teacher let the students translate the text into Japanese. Is she effective?”) was the most difficult.



Note. Each '#' is 2 raters.  
 Figure 14. Wright map for *Practice and Habit Formation*.

The Rasch model accounted for 46.4% of the variance in the observations.

The unexplained variance in the first contrast was 15.5%, and the eigenvalue was

1.4. The eigenvalues for the remaining contrasts were equal to or less than 1.4.

Because of the relatively large amount of variance accounted for by the Rasch model and the small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

A one-way ANOVA was conducted to determine if there were any significant differences in the ratings of the three groups. The independent variable, the raters, included the three groups: the student teachers, the college teachers, and the junior high school teachers. The dependent variable was the ratings for the *Practice and Habit Formation* construct. The ANOVA was significant,  $F(2, 101) = 10.85, p = .00$ , partial  $\eta^2 = .17$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the group sizes differed, and the homogeneity of variances test was not significant, the Games-Howell procedure was selected. There was no significant difference in the means between the student teacher group and the college teachers ( $p = .89$ ); however, there was a statistically significant difference between the student teachers and junior high school teachers ( $p < .001$ ) and between the college teacher and junior high school teachers ( $p = .05$ ).

*Construct 3: Communicative Practice*

The construct underlying questions 23, 32, 34, and 38 was called *Communicative Practice*. An example question was, “Overall, is the game on the screen planned by the student teacher effective?”

When the category functioning was checked, it was found that category 1 was used only 2 times (0%), so categories 1 and 2 were combined. In addition, category 5 was used only 21 times (5%), so categories 4 and 5 were combined. The resulting category structure met the criteria proposed by Linacre (1999) (see Table 43).

Table 43. *Rating Scale Functioning for Communicative Practice*

Category	Count (%)	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1 Not effective	43 (13)	.93	.92	NONE	(-9.17)
3 Neutral	112 (35)	1.03	1.06	-3.57	.00
4 Effective	169 (52)	1.02	1.01	3.57	(9.17)

The Rasch model explained 36.8% of the variance with an eigenvalue of 2.3. Unexplained variance in the first contrast was 23.7%, and the eigenvalue was 1.5. The eigenvalues for the remaining contrasts were equal to or less than 1.4. Because of the relatively large amount of variance accounted for by the Rasch model and the

small eigenvalues associated with the residuals, it was concluded that the construct was fundamentally unidimensional.

As shown in Table 45, all of the part-measure correlations were sufficiently strong, as they ranged from .60-.75, fit to the Rasch model was good, and the item difficulty estimates varied to a reasonable degree.

Table 44. *Rasch Descriptive Statistics for the Communicative Practice Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure Correlation
Question 23	48.0	1.0	.91	-.5	1.06	.4	.61
Question 32	52.7	.8	1.05	.4	1.02	.2	.66
Question 34	52.8	.8	.85	-1.1	.82	-1.3	.75
Question 38	46.5	1.0	1.21	1.0	1.10	.5	.60

*Note.* Pt-measure Correlation = Part-measure correlation; All statistics are based on Rasch CHIPS.

The mean of the student teacher group (50.65) was higher than those of the college teacher group (47.69) and the junior high school teachers' group (46.64). The 95% confidence intervals indicate that the mean of the student teacher group [56.60, 59.87] was significantly higher than the mean of the junior high school teacher group [52.21, 56.88].

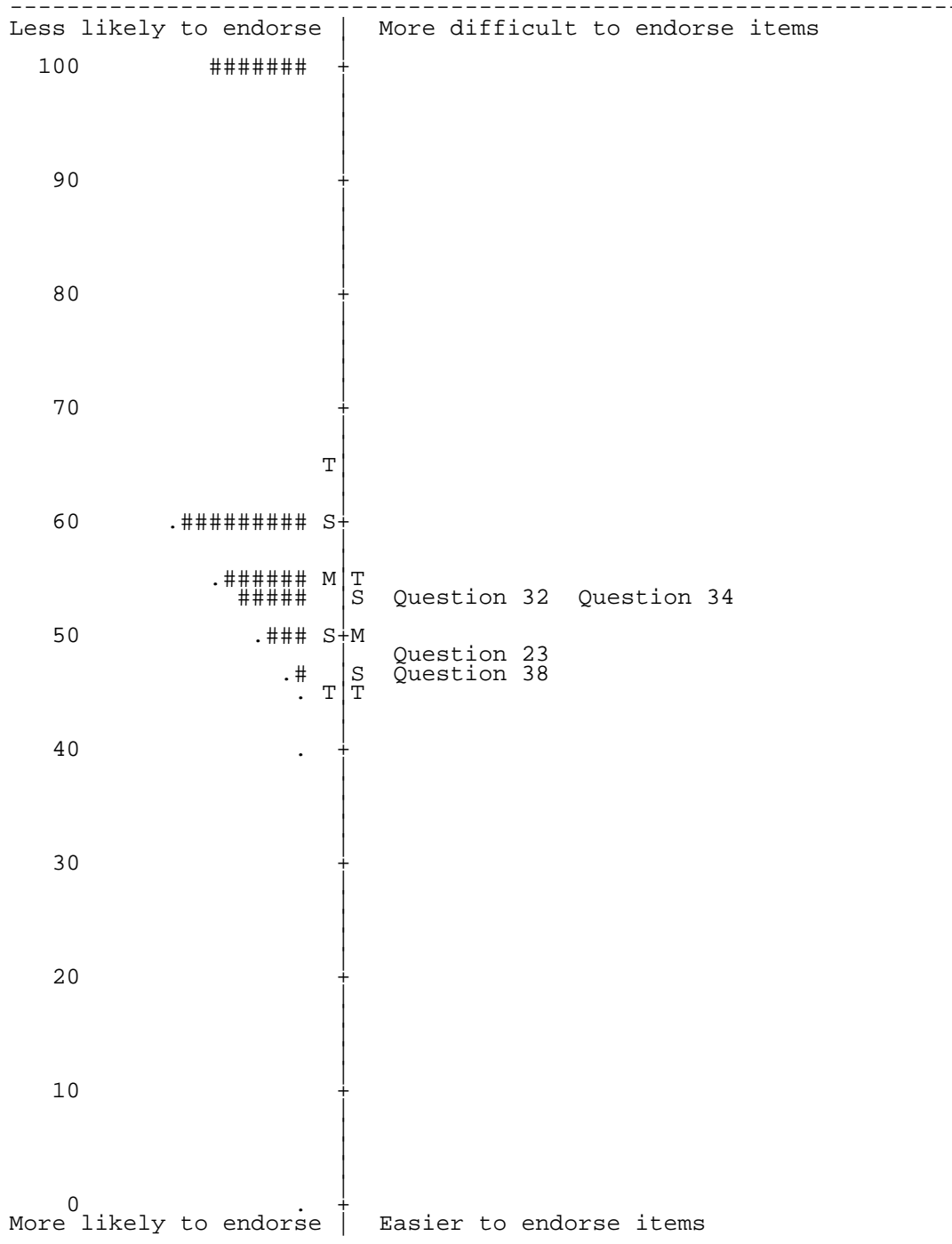
Table 45. *Descriptive Statistics for Communicative Practice*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	58.24	54.17	54.54
<i>SE</i>	0.81	1.97	1.14
95% CI Lower Bound	56.60	50.04	52.21
95% CI Upper Bound	59.87	58.30	56.88
<i>SD</i>	6.15	8.57	6.02
Skewness	-0.54	-1.23	0.01
<i>SES</i>	0.32	0.52	0.44
Kurtosis	0.11	1.76	0.04
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on Rasch person measure estimates.

The Rasch item reliability (separation) was good at .89 (2.80), but the Rasch person reliability (separation) was low at .06 (.25); this indicated that the person estimates were not well separated in relation to the standard errors.

The person mean (*SD*) of 54.78 (4.73) was somewhat higher than the item mean (*SD*) of 50.00 (2.80), indicating that the items were somewhat easy to endorse. Question 38 (“The student teacher on the screen encourages the students to recite what they have memorized by giving hints, paraphrasing, correcting, completing the sentences for the. Is she effective?”) was the easiest item to endorse (difficulty estimate = 46.5), and question 34 (“Is the student teacher’s reading aloud instruction effective?”) was the most difficult (difficulty estimate = 52.8).



Note. Each '#' is 2 persons.  
 Figure 15. Wright map for *Communicative Practice*.

A one-way ANOVA was conducted to determine if there is any significant difference in the ratings of the three groups. The independent variable, the raters, included the three groups: the student teachers, the college teachers, and the junior high school teacher group. The dependent variable was the ratings for the *Communicative Practice* construct. The ANOVA was significant,  $F(2, 101) = 4.37$ ,  $p = .015$ , partial  $\eta^2 = .08$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the group sizes were different but the test of homogeneity of variances was not significant, the Games-Howell procedure was used. There was a significant difference in the means between the student teacher group and the college teachers ( $p = .022$ ) and the student teacher group and the junior high school teachers ( $p = .017$ ). However, there was no significant difference between the college teachers and the junior high school teachers ( $p = .85$ ).

### *Differently Worded Likert Scales*

#### *Question 18: The Authenticity of the Language Game*

Question 18 asked the raters if the language game on the screen was authentic. They selected one of the following ratings: 1 = *Very authentic*, 2 = *Somewhat*

*authentic*, 3 = *Neutral*, 4 = *Not very authentic*, or 5 = *Not at all authentic*. The raw scores were used in this analysis because only one item was analyzed.

The mean of the college teacher group (3.58) was higher than those of the student teacher group (2.70) and the junior high school teacher group (2.82). The 95% confidence intervals of the college teacher group [3.25, 3.91] showed no overlap with those of the student teacher group [2.50, 2.90] and junior high school teacher group [2.49, 3.16]; thus, the difference in means is significant.

A one-way ANOVA was conducted. The independent variable was the three groups of raters (i.e., the student teachers, the college teachers, and the junior high school teachers), and the dependent variable was the ratings of the three groups for Question 18. The ANOVA was significant,  $F(2, 101) = 9.30$ ,  $p = .00$ , partial  $\eta^2 = .16$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. The Games-Howell procedure was used because the group sizes were different. There was a significant difference in the means between the student teachers and the college teachers ( $p = .00$ ) and also between the college teachers and the junior high school teachers ( $p = .00$ ), but no significant difference

Table 46. *Descriptive Statistics for Question 18*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	2.70	3.58	2.82
<i>SE</i>	0.10	0.16	0.16
95% CI Lower Bound	2.50	3.25	2.49
95% CI Upper Bound	2.90	3.91	3.16
<i>SD</i>	0.76	0.69	0.86
Skewness	-0.99	-0.31	-1.12
<i>SES</i>	0.32	0.52	0.44
Kurtosis	2.22	0.27	2.99
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on raw scores

was found between the student teachers and the practicum supervisors ( $p = .81$ ).

The  $\eta^2$  value was large at .16. Thus, the college teacher group gave significantly more severe ratings than the other groups. The college teachers tend not to think that the language game on the screen was authentic in comparison with the other groups.

#### *Question 20: Student Interest in the Language Game*

Question 20 asked if the game shown on the videotape can be interesting to students. The raters selected from the following responses: 1 = *I strongly disagree*, 2 = *I disagree*, 3 = *Neutral*, 4 = *I agree*, and 5 = *I strongly agree*. The raw scores were used for the analysis.

The mean of the student teacher group (2.16) was lower than those of the college teacher group (2.58) and the junior high school teacher group (2.82). The confidence intervals of the student teacher group [1.90, 2.41] slightly overlapped those of the college teacher group [2.18, 2.98] and were completely separated from those of the junior high school teachers [2.47, 3.17].

A one-way ANOVA was conducted. The independent variable was the three groups of raters: the student teachers, the college teachers, and the junior high school teachers. The dependent variable was the ratings of the three groups. The ANOVA was significant,  $F(2, 101) = 5.21, p = .007$ , partial  $\eta^2 = .09$ .

Table 47. *Descriptive Statistics for Question 20*

	Student Teachers	College Teachers	Junior High School Teachers
<i>M</i>	2.16	2.58	2.82
<i>SE</i>	0.13	0.19	0.17
95% CI Lower Bound	1.90	2.18	2.47
95% CI Upper Bound	2.41	2.98	3.17
<i>SD</i>	0.96	0.84	0.90
Skewness	0.43	0.36	0.05
<i>SES</i>	0.32	0.52	0.44
Kurtosis	-0.16	-0.48	-1.19
<i>SEK</i>	0.62	1.01	0.86

*Note.* All statistics are based on raw scores.

Follow-up tests were conducted to evaluate pairwise differences among the means. The Games-Howell procedure was used because the group sizes were different. There was a significant difference in the means between the student teachers and the junior high school teachers ( $p = .008$ ), but no significant differences between the college teachers ( $p = .44$ ) and the junior high school teachers, or between the student teachers and the college teachers ( $p = .44$ ). The  $\eta^2$  value was medium at .09. The student teacher group did not think that the game was as interesting to the junior high school students as the other groups.

*Question 21: Appropriacy of Language Activities*

Question 21 asked the raters which type of language activities is most appropriate to junior high school students? The raters were asked to choose one of the following alternatives:

1. Performing memorized dialogues
2. Contextualized drills in which everything is controlled by the teacher
3. A language game in which students use a new sentence pattern. They can choose any words they like.
4. Cued dialogues

## 5. Improvisation

Half of the student teachers selected category 3 (29, 50.88%), followed by category 4 (15, 26.32%). Over 60% of the college teachers selected category 3 (12, 63.16%), followed by category 4 (5, 26.32%). 11 (39.29%) of the junior high school teachers selected category 3, followed by category 4 (10, 35.71%). Many more of the student teachers and college teachers chose category 3 rather than category 4. The junior high school teachers selected category 3 almost the same number of times as category 4.

Table 48. *Crosstabulation Results for Question 21*

	Student Teachers	College Teachers	Junior High School Teachers	Total
Category1	1 (1.8%)	0 (0%)	3 (10.7%)	4 (3.9%)
Category2	2 (3.5%)	1 (5.3%)	1 (3.6%)	4 (3.9%)
Category3	29 (50.9%)	12 (63.2%)	11 (39.3%)	52 (50%)
Category4	15 (26.3%)	5 (26.3%)	10 (35.7%)	30 (28.9%)
Category5	9 (15.8%)	1 (5.3%)	2 (7.1)	12 (11.5%)
Total	56 (98.2%)	19 (94.7%)	27 (96.4%)	102 (98.1%)

*Note.* One student teacher and one junior high school teacher did not answer question 21.

The raw scores were used for analysis. As the data were categorical, a chi-square test was run. The results were not significant:  $\chi^2 = 9.65$ ,  $df = 10$ ,  $p = .47$ .

*Question 22: Type of Language Activity*

Question 22 asked the raters which type of language activities they think the game on the screen is. The raters were asked to choose from the following choices:

1. Performing memorized dialogues; 2. Contextualized drills in which everything is controlled by the teacher; 3. A language game in which students are supposed to use a new sentence pattern (they can choose the words); 4. Cued dialogues, or 5.

Improvisation. Category 3 were selected most frequently by all the groups [student teachers = 30 (52.6%), college teachers = 12 (63.2%), junior high school teachers =

15 (53.6%)], followed by category 4 [20 (35.1%), 4 (21.1%), 8 (28.6%),

respectively]. As the data were categorical, a chi-square test was run; the results

were not significant,  $\chi^2 = 6.52$ ,  $df = 10$ ,  $p = .77$ .

Table 49. *Crosstabulation Results for Question 22*

	Student Teachers	College Teachers	Junior High School Teachers	Total
Category 1	1 (1.8%)	2 (10.5%)	3 (10.7%)	6 (5.8%)
Category 2	4 (7.0%)	1 (5.3%)	2 (7.1%)	7 (6.7%)
Category 3	30 (52.6%)	12 (63.2%)	15 (53.6%)	57 (54.8%)
Category 4	20 (35.1%)	4 (21.1%)	8 (28.6%)	32 (30.8%)
Category 5	1 (1.8%)	0 (0%)	0 (0%)	1 (1.0%)
Total	56 (98.2%)	19 (94.7%)	27 (96.4%)	102 (98.1%)

*Note.* One student teacher and one junior high school teacher did not answer question 22.

## CHAPTER 7

### DISCUSSION

In this chapter, I first discuss research question 2. The discussion of research question 1 is taken up after the discussion of research questions 3 and 4 because the answers to those two questions have a strong bearing on the discussion of research question 1.

#### Research Question 2: The Teaching Skills Questionnaire

Research question 2 asked to what degree the student teachers, college teachers, and junior high school teachers' ratings on the *Teaching Skills Questionnaire* differed. As no previous researchers have compared the perceptions of student teachers, junior high school teachers, and college teachers concerning the student teachers' teaching skills in a language teaching practicum, it is not possible to compare the results with previous research.

### *Construct 1: Teaching Mainly in Japanese*

The construct underlying questions 1-5 was *Teaching Mainly in Japanese*. A representative item is “To explain a new structure in Japanese.” The mean Rasch person ability estimates (CHIPS) for the college teacher group (61.16) and the junior high school teacher group (61.15) were considerably higher than the mean (51.97) of the student teacher group. The higher means indicated that the college teachers and junior high school teachers thought that items 1-5 were easier to teach than the student teachers. When the group means were compared, a one-way ANOVA indicated a statistically significant difference overall,  $F(2, 99) = 16.01, p = .00$ , partial  $\eta^2 = .25$ . Follow-up tests showed that the significant differences were between the mean ratings between the student teachers and the college teachers ( $p = .001$ ) and between the student teachers and the junior high school teachers ( $p = .001$ ); no statistically significant difference was found between the college teachers and the junior high school teachers ( $p = 1.00$ ).

The similar perceptions of the college teachers and junior high school teachers could have occurred because of their many similarities: They are similar ages, both groups are mature adults, they have considerable teaching experience, and they have similar levels of English knowledge. Also, the college teachers were

involved in the teacher licensure program, so they were teaching courses in teaching methodology to the college students. Much of the content of these courses was similar to the pedagogical approach adopted by the junior high school English teachers.

In contrast to the college teacher and junior high school teacher groups, the student teachers had little experience teaching at the junior high school level, and this lack of teaching experience might have accounted for some of the observed differences, as it might have led them to perceive the five items assessed on this part of the questionnaire as being difficult to teach. The student teachers were differentiated from the older teacher groups because of their limited teaching experience and lower English proficiency. Five student teachers had a TOEIC score under 400, 22 of the student teachers had a TOEIC score greater than 500, and just four of the 22 students had obtained a score above 600 when they participated in their practicum. Junior high school teachers expect student teachers to have a TOEIC score greater than 700 before participating in the practicum. In addition, the student teachers, as young adults, might have found it difficult to manage the younger junior high school adolescents. The fact that the student teachers were approximately 21 years old might have influenced some of the junior high school

students, who saw them not as teachers but more like older friends. Because some of the junior high school students developed a relatively close and friendly relationship with the student teachers and because this might have reduced the student teachers' authority in the eyes of some of the students, the student teachers might not have been able to manage the class effectively at times. For example, two student teachers said that they were unable to tell the junior high school students to be quiet when they were noisy, because they were afraid that their good relationship with those students would have been compromised. Thus, even in cases when the student teachers' skills were adequate, they might have found teaching difficult at times because of their overly friendly relationship with the junior high school students.

Junior high school students are usually friendly and cooperative with student teachers for at least three reasons. The first reason is that the student teachers have had no contact with junior high school students, so they have no information about them; this lack of knowledge seems to allow the student teachers to communicate freely and in an unbiased way with the junior high school students. In contrast, some junior high school teachers felt that some of the junior high school students are someone to watch out for because they have broken the school

rules previously. Some of the critical comments made by the junior high school teachers have also hurt the feelings of some of the junior high school students, so these students feel somewhat uneasy with these English teachers, but not with the student teachers. The second reason concerns the fact that many junior high school students do not have siblings, as they are the only child in their family. Having contact with young teachers is a new experience for them, and they often feel closer to them than to their junior high school teachers, who tend to be considerably older. The third reason is that the student teachers are not particularly skilled at teaching English, and the junior high school students react to this by cooperating with the student teachers as they struggle to conduct the class. Junior high school students look for sensitive teachers who can understand them and can teach English in a way that leads them to come to like English. Even when talented junior high school students who are unusually good at English think that their progress in English class is slowed by the student teachers, they do not generally show negative attitudes in class.

### *The Empirical Item Hierarchy for Teaching Mainly in Japanese*

The items making up the *Teaching Mainly in Japanese* construct were expected to be ordered from item 1 (easiest to do) to item 5 (most difficult to do). The Rasch item reliability estimate was 1.00 (Rasch item separation = 15.18), which indicated great differences among the item difficulty estimates in relation to their standard errors. As predicted, item 1 (*To greet in English*) was the easiest task for the participants to engage in (Rasch item difficulty estimate = 46.2), and items 2 (*To answer it when a student asks about the Japanese meaning of an English word in the textbook*) (Rasch item difficulty estimate = 49.4) and 4 (*To explain a new structure in Japanese*) (Rasch item difficulty estimate = 49.4) were the next easiest. Items 3 (*To translate text in the textbook into Japanese*) and 5 (*To give background knowledge of the text in the textbook*) were more difficult.

Item 4 (*To explain a new structure in Japanese*) was expected to be more difficult than Item 3 (*To translate text in the textbook into Japanese*) because the student teachers had considerable experience translating English texts into Japanese when they were junior and senior high school students; however, this item was somewhat difficult for them because of their limited experience providing grammatical explanations of the sentences they translated. A second issue affecting

the difficulty of item 3 concerned the scores awarded by each group. The student teachers, college teachers, and junior high school teachers had raw score means of 3.13, 1.84, and 2.36, respectively for item 3, and raw score means of 3.00, 1.79, and 1.39, respectively for item 4; thus, all three groups of raters assessed item 4 as being easier, with the greatest difference being displayed by the junior high school teachers.

Table 50. *Mean Scores of the Three Groups for Items 3 and 4*

	Item 3	Item 4
Student teacher	3.13	3.00
College teacher	1.84	1.79
Junior high school teacher	2.36	1.39

*Note.* Means were calculated using raw scores. Higher scores indicate that the raters thought an item was more difficult to teach.

One junior high school teacher stated that the variance in academic ability among the Japanese students was quite wide; according to the teacher, some of the junior high school students found it difficult to understand the contents of the text even in Japanese. For this reason, very skillful Japanese translations were necessary if the lower proficiency junior high school students were to comprehend the meaning and cultural aspects of the English text. This variance in academic ability among the junior high school students might have influenced the perceived

difficulty of item 3 because it was difficult for the student teachers to provide translations that satisfied the junior high school students or were comprehensible to them. The junior high school students generally value the ability to produce accurate translations because they are seen as the key to succeeding on English tests and the ability to translate skillfully is viewed as one of primary goals of studying English.

Overall, the three groups found the skills measured by items 1-5 (*Teaching Mainly in Japanese*) easiest for the student teachers to perform. This was likely caused by the fact that these skills were performed in Japanese, rather than English.

#### *Construct 2: Teaching Using Easy English*

The construct underlying questions 6-10 was *Teaching Using Easy English*. A representative item is “To instruct students in classroom English.” The mean Rasch person ability estimates (CHIPs) of the college teachers and the junior high school teachers were similar at 58.81 and 59.50, respectively. Both means were higher than the mean of the student teachers (46.33). An analysis of the group means with Welch’s  $F$  was significant,  $F(df\ 2, 35.73) = 23.29, p = .00$ , partial  $\eta^2 = .31$ , and follow-up tests indicated that the significant differences in means were

between the student teachers and the college teachers ( $p = .001$ ) and between the student teachers and the junior high school teachers ( $p = .000$ ). No statistically significance was found between the college teachers and the junior high school teachers ( $p = .977$ ).

Once again, the college teachers and the junior high school teachers' ratings were similar. This might have occurred because of the commonalities noted above between these two groups: their ages, teaching experiences, understanding of teaching methodology, and level of English proficiency. In contrast, the participants in the student teacher group, who were much younger, with less teaching experience, and less knowledge of English, awarded significantly different ratings than the two groups of older, more experienced, and more knowledgeable teachers.

#### *The Empirical Item Hierarchy for Teaching Using Easy English*

For the items used to measure *Teaching Using Easy English*, it was predicted that the lower numbered items would be easier based on the results of the pilot study; thus, the order of difficulty should have gone from item 6 (easiest) to item 10 (more difficult). While this prediction generally held true, one exception was

that item 9 (*To instruct students in classroom English*) was easier than item 8 (*To answer grammatical questions*).

Table 51. *Mean Scores of the Three Groups for Items 8 and 9*

	Item 8	Item 9
Student Teachers	4.44	3.87
College Teachers	2.74	2.53
Junior High School Teachers	2.36	2.43
Overall Mean	3.55	3.23

*Note.* Means were calculated using raw scores. Higher scores indicate that the raters thought an item was more difficult to teach.

The raw scores indicated that the junior high school teachers saw items 8 and 9 as being approximately equally difficult ( $M = 2.36, 2.43$ , respectively); however, the student teacher group and the college teacher group thought that item 9 (*To instruct students in Classroom English*) (3.87, 2.53, respectively) was easier than item 8 (*To answer grammatical questions*) (4.44, 2.74, respectively), and the largest difference was displayed by the student teachers. The higher ratings for item 9 might have occurred because of the student teachers' limited experience explaining English grammar, their difficulties providing examples beyond those they had prepared before class, and the simplicity and repetitiveness of classroom English (e.g., *Stand up, Go back to your seat*).

Problems explaining grammar points might have occurred for two reasons.

First, five of the student teachers prepared for their teaching lessons by considering how to explain the new target grammar points in the lesson. However, because they had not fully internalized their understanding of those grammar points, they tried to explain the points by quoting explanations they had found in grammar books. The junior high school teachers who supervised these student teachers said that when the student teachers' explained these grammar items, it sounded as if they were reading from a grammar book. Second, the student teachers taught both new grammar items and grammar items that the junior high school students had studied previously but which they had not yet understood very well. This led to the junior high school students asking questions about points that the student teachers were unprepared to answer, so they either made a weak attempt to answer the questions or asked the junior high school students to wait until the following day for the answer, as this gave them the opportunity to consult a grammar book for the answer. Because of these difficulties, the student teachers stated that instructing the students in English was easier than explaining grammar.

### *Construct 3: Teaching Using Difficult English*

The construct underlying questions 11-15 was *Teaching Using Difficult English*. A representative item is “To prepare for the next team-teaching class by discussing the class with an ELT in English.” The mean Rasch ability estimates (CHIPS) for the college teachers and the junior high school teachers were 59.93 and 52.30, respectively. These group means were considerably higher than the mean (44.98) of the student teachers. A comparison of the mean differences using Welch’s  $F$  was significant,  $F(df\ 2, 37.279) = 19.528, p = .00$ , partial  $\eta^2 = .32$ , and follow-up tests identified a significant difference between the student teachers and the college teachers ( $p = .001$ ) and between the student teachers and the junior high school teachers ( $p = .000$ ). No significance difference was found between the college teachers and the junior high school teachers ( $p = .249$ ). As discussed above, the many similarities between the two older teacher groups, such as their age, amount of teaching experience, and similar levels of English proficiency, likely contributed to their similar perceptions of the teaching on the videotapes.

### *The Empirical Item Hierarchy for Teaching Using Difficult English*

Based on the results of the pilot study, it was predicted that the lower numbered items would be easier; thus, the item difficulty hierarchy was expected to go from item 11 (easiest) to item 15 (most difficult); however, the data indicated that item 11 (*To give good examples of the new structure*) was a little more difficult than item 12 (*To ask questions in English about text and let students try to answer in English*) and item 13 (*To make an authentic and interesting language activity such as information gap for your class*). The two most difficult items, item 14 (*To introduce context in your English words*) and item 15 (*To prepare for the next team-teaching class by discussing the class with an ELT in English*) were in the hypothesized order.

One reason that item 11 was more difficult than predicted might be that many of the student teachers taught first-year junior high school students. The reading texts and new sentence patterns introduced in the first half of the first year of junior high school are very simple, so it was not difficult for the student teachers to ask questions about these materials in English and have the junior high school students answer in simple one- or two-word responses that were based on information explicitly stated in the text. The reading texts were written using high frequency

vocabulary, no abstract concepts were introduced, and no inferential questions were asked. These factors made the texts both easy to teach and easy for the junior high school students to comprehend.

In regards to Item 13 concerning interactive activities (e.g., information-gap activities) for first-year students, the sentence patterns introduced in these tasks were simple, and considerably easier than those introduced to the second- and third-year students, whose tasks include more complicated syntactic structures such as relative clauses. The student teachers were expected to provide clear examples of the new structure that were closely related with the junior high school students' interests and everyday life and that therefore had relevance to them. Although the student teachers prepared examples beforehand, they had to quickly provide other examples if the junior high school students did not understand their prepared examples. The student teachers stated that giving good examples on the spot was difficult.

One junior high school teacher who commented on Item 15 had a good command of English, yet he thought that teaching new sentence patterns in English and providing clear examples of the pattern was fairly difficult because of the diverse English proficiency levels of the students in the typical junior high school

classroom. According to him, item 15 was not difficult in terms of the English necessary to carry out the task; what was difficult was communicating effectively with the lower proficiency students in the class so that they would be able to understand the point.

Overall, items 11-15 (*Teaching Using Difficult English*), measured skills that were seen by the raters as being the most difficult for the student teachers to do well. This is based on the assumption that teaching skills requiring greater English proficiency are more demanding and more difficult to carry out successfully (See Figure 16), but this does not always mean that teachers with better English proficiency can carry out these teaching skills more successfully. As one junior high school teacher commented in the last paragraph, they also need to make them understood to students with the lower English proficiency.

The college teachers' mean ratings of items 6-10 and items 11-15 were nearly identical, and they were relatively lenient in both cases. This might have occurred in part because their English proficiency is much higher than the proficiency level required to teach the skills measured by these items; as a result, they might not have differentiated among the items. It is also possible that they did not carefully consider the classroom situation and the students' diverse academic levels, the

presence of slow learners, learners with attention deficit problems, and students who were not emotionally stable. All of these issues might have been salient for the junior high school teachers and student teachers because they engaged in frequent face-to-face interaction with the junior high school students.

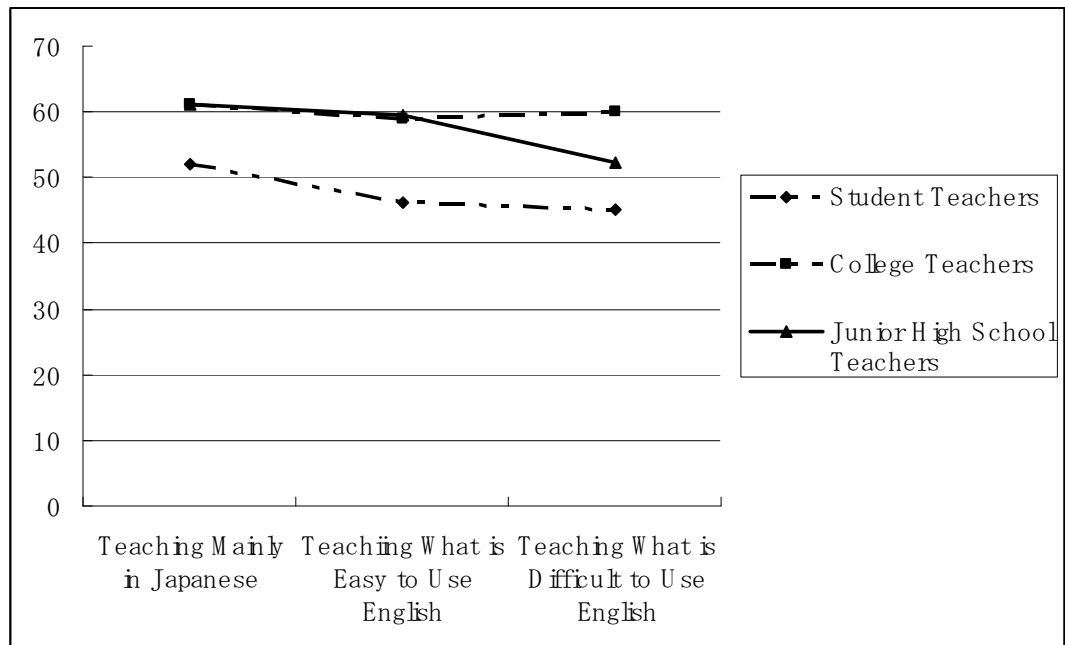


Figure 16. The mean Rasch person ability estimates (CHIPS) of each teacher group for the three constructs.

### Research Question 3: Differential Use of the Revised COLT

The third research question asked to what degree the student teachers, practicum supervisors, and college teachers' ratings of student teachers' teaching performances differed when they used the *Revised COLT*. This question was answered by analyzing the three groups' ratings for the *Student Teachers' Teaching on the Video*. No difference was found among the three groups for student teachers 1, 2, 3, 4, and 5. Figure 17 shows the three groups' ratings for student teachers 1-5.

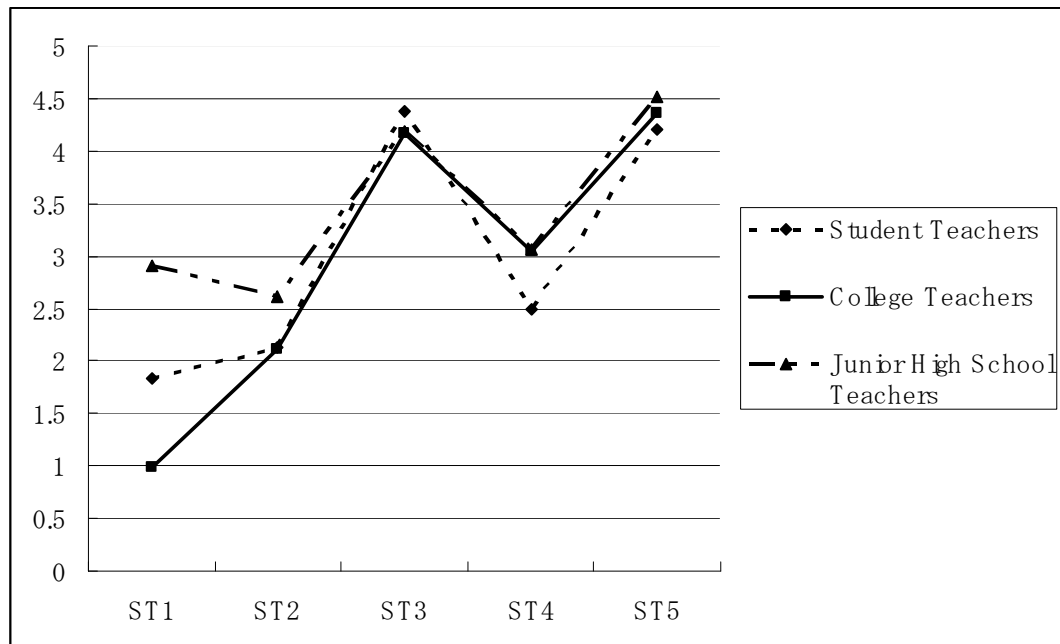


Figure 17. The three groups' mean ratings (Rasch logits) for student teachers 1-5.

The participants awarded ratings of 0 (*the student teacher on the video did not engage in the behavior*), 1 (*the teacher engaged in the behavior to a degree*), or 2 (*the teacher clearly engaged in the target behavior*) for each of the categories they rated (The participants did not have to write anything in some categories. I filled in those parts because of the educational context in this study. Boxes filled with a circle are shown in Appendix H.). These data were analyzed with the multi-

faceted Rasch model. The mean ratings (Rasch logits) for each group are shown in Table 52.

Table 52. *The Mean Ratings (Rasch Logits) for the Three Groups for Students 1-5*

Observed Teacher	Student Teacher Ratings	College Teacher Ratings	Junior High School Teacher Ratings
ST1	1.84	0.98	2.90
ST2	2.12	2.11	2.62
ST3	4.37	4.17	4.18
ST4	2.50	3.05	3.06
ST5	4.21	4.36	4.51

The ratings were analyzed for group differences with a one-way ANOVA.

None of the ANOVAs reached statistical significance; however, the junior high school teacher group produced the highest mean on four out of the five occasions, indicating that they were relatively lenient. The one exception was for student teacher 3; in that case, the student teacher group awarded the highest overall ratings. ST1's mean ratings from the student teacher group, the college teacher group, and the junior high school group were 1.84, 0.98, 2.90, respectively. Though there were not any statistically significant differences among the three groups ( $p = .51$ ), the differences were still noteworthy in that the junior high school teachers gave the highest ratings to Student teacher 1 for *Teaching review and new sentence pattern*.

This indicates that they noticed that Student teacher 1 engaged frequently in the behaviors described in *L2 use* and its related activity categories more than the other rater groups.

There are two possible interpretations of this finding. One is that the junior high school teacher group were more aware of the student teachers' or the students' behavior than the other two groups, and the other is that the junior high school teacher group noticed and regarded subtle behaviors as important and awarded higher scores based on those subtle behaviors, i.e., they awarded 1s or 2s where raters in other groups awarded 0s or 1s. Both interpretations suggest that the junior high school teachers might have been more sensitive to the behavior of the student teachers. For example, two of the junior high school teachers involved in this study stated that they noticed that the student teachers sometimes missed hints from the junior high school students that they wanted to communicate with the student teachers. Three more junior high school teachers also noticed that the student teachers tended to focus on academically capable students and assertive students and pay less attention to students with average academic ability and quiet students. Several student teachers confirmed this perception in their teaching diaries, which they submitted to their junior high school supervisors and then to the college with

their practicum supervisors' signatures after completing the practicum. This behavior cannot be confirmed by viewing the videotaped lessons because the video camera was focused on the student teachers; thus, the observers were unable to see the junior high school students' behavior on the videotape.

#### Research Question 4: The Student Teachers' Teaching on the Video

The fourth research question asked to what degree the student teachers, college teachers, and junior high school teachers' responses differed when they used the *Questions Concerning the Student Teachers' Teaching on the Video*. First, the answers to the 22 questions concerning teaching methods are discussed. These 22 questions measured four constructs: *Listening and Japanese Use*, *Prereading*, *Using English*, and *Memorization*. The results were used to investigate differences among the three rater groups' perceptions of these teaching methods.

#### *Constructs Related to the Usefulness of the Instruction*

##### *Listening and Japanese Use*

The mean Rasch person ability estimates (CHIPS) of the student teacher group, the college teacher group, and the junior high school teacher group were

57.44, 57.33, and 54.56, respectively; no statistically significant differences were identified,  $F(2, 101) = 1.39, p = .253$ . Thus, overall, the views of the three groups were similar where *Listening and Japanese Use* is concerned. This is a positive outcome because it indicates that what the student teachers studied, what the college teachers taught in the teacher licensure programs, and what the junior high school teachers practiced in the junior high schools resulted in similar perceptions.

The item difficulty hierarchy was, from the easiest to the most difficult item, Question 7 (*Is it useful to explain about English grammar in Japanese?*), Question 30 (*Is it useful to let students translate the text into Japanese?*), Question 13 (*Is it useful to ask questions in Japanese about the text?*), Question 39 (*Is it useful for students to translate one Japanese sentence into English using the new sentence pattern?*), and Question 5 (*In these two scenes, what the students did was just listening except on some occasions. Is this useful for review?*). Question 5 was more difficult to endorse than the other questions by approximately 2 CHIPS. What made Question 5 different from the other questions was the use of English. If the student teacher had explained the contents and grammar of the listening passage in Japanese, Question 5 might have been easier to endorse. Most of the raters chose *Not useful* or *Not very useful* for question 5, which suggests that the raters were not

satisfied with the student teachers' English, which was clearly weak, even though they were supposed to rate the usefulness of the teaching methods.

### *English Use*

The mean Rasch estimates (CHIPS) of the student teacher group, the college teacher group, and the junior high school group for *English Use* were 51.81, 49.50, and 52.95, respectively. Welch's  $F$  was conducted to determine if there were any significant differences in the mean ratings of the three groups. The result was not significant,  $F(2, 53.95) = 1.62, p = .207$ , indicating that their perceptions of the questions used to measure this construct were similar.

The item difficulty hierarchy was, from the easiest to the most difficult, Question 35, 33, 19, 11, 3, 1: Question 35 (*This is not what the student teacher on the screen is doing. Is it useful to give students background knowledge of the text and/ or activate their background knowledge before going into the text?*), Question 33 (*Is it useful to let students practice reading aloud the text?*), Question 19 (*Is it useful if a language game is interesting to students?*), Question 11 (*Generally speaking, in order to learn a new sentence pattern, is it useful for students to practice such language activities as information gap/ interview games?*), Question

3 (*Is it useful if a teacher introduces a new sentence pattern in English trying to lead students to find out a new rule?*), and Question 1 (*Is it useful if a teacher reviews the class in English trying not to use Japanese?*). The difference in the item difficulty estimates between Question 1 and 19 was 1.4 CHIPS. Questions 1 and 3 were similar in that they both involved the exclusive use of English. Questions concerning the use of only English were somewhat difficult for the raters to endorse. The raters provided several reasons in the memos they wrote in the margin of the questionnaire answer sheet. First, many of them believed that the student teachers' exclusive use of English was difficult for the junior high school students to comprehend and that the use of Japanese would help the junior high school students comprehend more fully. Second, they felt that the student teachers' English was not good enough to handle the pedagogical tasks. I discuss this issue at the end of the chapter in relation to other findings. The five student teachers' had low-intermediate to intermediate English proficiency: One had passed the 2<sup>nd</sup> level of the EIKEN, a commonly used English proficiency test is Japan Society for Testing English Proficiency in Japan. This level is approximately equivalent to a 500-550 TOEIC score. Another of the teachers had a 600 TOEIC score, and the others had TOEIC scores under 450.

### *Prereading*

The mean Rasch estimates (CHIPS) of the student teachers, college teachers and junior high school teachers were 50.59, 49.10, and 48.94, respectively. The results of a one-way ANOVA were not significant,  $F(2, 101) = .48, p = .62$ , again indicating that their perceptions of the questions used to measure this construct were similar.

The item difficulty hierarchy was, from the most difficult to the easiest, Question 9 (61.80), 36 (45.60), and 24 (42.60): Question 9 (*Except for some occasions, the students just listened. Is this useful for introducing the new sentence pattern?*), 24 (*Is it useful to introduce new words with activities such as flash cards, checking the pronunciation, and checking word meanings, before going into a text?*), and 36 (*This is not what the student teacher on the screen is doing. Is it useful to let students grasp the points of the text before reading intensively?*). Question 9 differed from the other questions because the activity related with Question 9 required the use of English. It seems to be difficult for the raters to endorse items in which Japanese is not used.

### *Memorization*

The mean Rasch estimates of the student teachers (44.57) and college teachers (44.01) were lower than that of the junior high school teachers (49.84). A one-way ANOVA indicated that there were no significant differences among the groups,  $F(2,101) = .986, p = .377$ ; thus, the three groups' perceptions of the questions used to measure this construct were similar.

Item 9 (*In order to learn a new sentence pattern, is this useful for students to repeat the new sentence pattern after the teacher?*) was considerably more difficult to endorse than item 37 (*Is it useful to let students memorize the text as part of post reading activities?*). The difference between these two items was 1.54 CHIPS. The standard deviations of the three groups were large (student teachers = 16.80; college teachers = 19.51; junior high school teachers = 17.11), which indicated that the participants in each group had widely differing views of these two items that are related to behaviorism, i.e., repetition and memorization. Junior high school English teachers frequently use both of these. This variance is caused by at least two factors. First, the raters formed a continuum between persons who strongly support behaviorism and those who do not. Second, the student teachers' use of repetition and memorization on the video was not carried out skillfully and this

likely influenced the ratings. In general, Japanese teachers' ideas about teaching English in junior high school, senior high school, and college/ university vary widely.

The rater groups had difficulty endorsing item 9 because the student teacher simply had the students repeat the sentence monotonously and did not create any variation in the task by including activities such as substitution drills and/or conversion drills. She also had the students repeat at a slow tempo. For these reasons, even the raters who supported behaviorism might have rated her performance severely.

#### *Constructs Related to the Effectiveness of the Instruction*

This section was made up of 16 questions concerning the effectiveness of the student teachers' teaching. Three constructs were identified in this part of the instrument, *Explanation and Understanding*, *Grammar-translation and Habit-formation*, and *Communicative Practice*.

### *Explanation and Understanding*

This construct was measured with items 2, 4, 12, 16, 27, 29, and 42. The mean Rasch estimates of the student teachers, college teachers, and junior high school teachers for *Explanation and Understanding* were 48.28, 44.89, and 44.78, respectively. A one-way ANOVA was significant,  $F(2, 101) = 5.56, p = .005$ , but the effect size, partial  $\eta^2 = .10$ , was low. Follow-up tests indicated that there was a statistically significant difference in the means between the student teachers and the college teachers ( $p = .036$ ), and between the student teachers and the junior high school teachers ( $p = .005$ ). The college teachers and the junior high school teachers were more severe in their ratings than the student teachers. This result indicates that the student teachers need to be trained to have better teaching skills and to be more skillful in their use of English. I discuss this issue at the end of the chapter.

The item difficulty hierarchy was, from the most difficult to the easiest, as follows: Question 2 (*The student teacher on the screen tries to review the class in English without using Japanese. Is she effective when her English and her teaching techniques are taken into consideration?*), Question 4 (*The student teacher on the screen tries to introduce a new sentence pattern in English by leading students to discover a new rule. Is she effective when her English and her teaching technique*

*are taken into consideration?), Question 27 (The student teacher introduces the text orally. Is she effective?), Question 12 (Is the interview game the student teacher on the screen is introducing effective?), Question 16 (Is the student teacher's instruction before the game and after the game effective?), Question 42 (The student teacher talks in Japanese why the text is worth reading as part of post reading activities. Is she effective?), and Question 29 (The student teacher let the students underline sentences which have the new sentence pattern so that they will be aware of the new sentence pattern in the text. Is she effective?).*

What made questions 2, 4, and 27 (Rasch item difficulty estimates = 57.4, 55.7, and 49.4, respectively) different from questions 12, 16, 42, and 29 (Rasch item difficulty estimates = 48.7, 48.3, 45.3, and 45.2, respectively) was that the student teachers used only English in the activities related to questions 2, 4, and 27; thus, the results indicate that the student teachers' English was perceived as not being good enough to handle those particular tasks. Many practicum supervisors stated that this group of student teachers should have had higher English proficiency. That said, the groups' views of English proficiency and its relationship with teaching communicatively differed in the sense that the student teacher group was more lenient. This might be related to the fact that the student teacher group

would find it more difficult to teach using difficult English due to their overall lower level of proficiency. Their lack of confidence in using English might have led them to be relatively lenient about the five teachers' use of English on the video.

### *Practice and Habit Formation*

The means of the student teacher group, the college teacher group, and the junior high school teacher group for *Practice and Habit Formation* were 48.61, 47.95, and 41.93, respectively. A one-way ANOVA indicated that the differences were significant,  $F(2, 101) = 10.85$ ,  $p = .00$ , partial  $\eta^2 = .17$ . Follow-up tests showed that there was a statistically significant difference between the student teachers and junior high school teachers ( $p < .001$ ) and between the college teachers and junior high school teachers ( $p = .05$ ). No significant difference was found between the student teachers and the college teachers. The junior high school teacher group was severe in their assessment of the student teachers' teaching on the video.

The item difficulty hierarchy was, from the easiest to the most difficult, Question 10 (47.9) (*The student teacher on the screen let her students repeat after her. Is she effective?*), Question 25 (49.2) (*Is the student teacher effective in*

*introduction of new words with flash cards checking the pronunciation, the meanings before going into a text?*), Question 8 (49.9) (*The student on the screen explains about new items of English grammar. Is she effective?*), Question 14 (50.5) (*The student teacher let students translate several Japanese sentence isolated from each other into English. Is she effective?*), and Question 40 (52.5) (*The student teacher let the students translate the text into Japanese. Is she effective?*); however, the difficulty estimates were not well separated.

In research question 2, the college teacher group and the junior high school group gave higher ratings to *translate text in the textbook into Japanese* than was predicted, probably because these two groups were more concerned with difficult aspects of translation such as cultural nuances. These higher ratings indicate that when the student teachers asked the junior high school students to translate the text into Japanese was not effective. On the video, one of the five student teachers was translating text in the textbook into Japanese. From the viewpoints of the college teachers and the junior high school teachers, her translation of the text was not entirely satisfactory for at least two possible reasons. One is that the student teacher's translation of the text did not demonstrate an awareness of the cultural aspects embedded in the text. On the video, the student teacher introduced a new

word, *sister*, and translated it in Japanese as *onna no kyoudai*. Strictly speaking, *sister* is singular, so it is *ane* (an older sister), or *imouto* (a younger sister) in Japanese. In the text the student teacher and the junior high school students were reading, the following sentence appeared: *She has a sister and she (the sister) works at a bookstore*. The student teacher translated this sentence into Japanese as: *Kanojo niha sisitah ga ite, sono sisutah wa honnya de hataraitairu*. She did not translate *sister* into Japanese and did not explain that in Japanese, people indicate whether a sister is an older sister (*ane*) or a younger sister (*imouto*), but native English speakers do not typically do so. The other issue is that she did not provide any explanations of the particular grammatical structures in the sentences; this would have allowed the students to understand why the sentences were translated into Japanese the way that they were.

The *Questions Concerning the Student Teachers' Teaching on the Video* was made up of 42 *questions*, most of which were paired. One was a question about methods related with the student teachers' teaching and the other concerned how well the student teachers put these methods into practice. In the ratings for the first group of the questions, no significant difference was found, so the significant one-way ANOVA results did not come from reasons concerning the teaching methods

and techniques mentioned in this group of questions; the differences apparently arose from how the student teachers used these teaching methods and techniques. In the practicum, the junior high school teachers advised the student teachers to have the junior high school students practice drills more quickly. The student teachers were surprised to see how rapidly the junior high school teachers had the students practice. Thus, it is possible that the junior high school teachers were dissatisfied with this aspect of the student teachers' performance.

#### *Communicative Practice*

The mean Rasch estimate (CHIPS) of the student teacher group (50.65) was higher than those of the college teacher group (47.69) and the junior high school teachers' group (46.64). The one-way ANOVA was significant,  $F(2, 101) = 4.37$ ,  $p = .015$ , but the effect size, partial  $\eta^2$ , was low at .08. There was a significant difference in the means between the student teacher group and the college teachers ( $p = .022$ ) and the student teacher group and the junior high school teachers (.017); there was no significant difference between the college teachers and the junior high school teachers ( $p = .849$ ). The college teachers and the junior high school teachers were more severe in rating the student teachers' teaching than the student teachers.

This indicated that these three groups should discuss their different perceptions and the reasons for them, and the student teachers should gain a better understanding of why the other two groups awarded lower ratings.

Some reasons for the difference were indicated by the item difficulty hierarchy that was, from the most difficult to the easiest, as follows: Question 34 (52.8) *Is the student teacher's reading aloud instruction effective?*; Question 32 (52.7) *Overall, is the student teacher effective in asking questions in either English or Japanese?*; Question 23 (48.0) *Overall, is the game on the screen planned by the student teacher effective?*; Question 38 (46.5) *The student teacher on the screen encourages the students to recite what they have memorized by giving hints, paraphrasing, correcting, completing the sentences. Is she effective?*

This indicated that the student teacher's model reading and her instruction were not good enough and that the student teacher needed to be trained to ask more questions in English or Japanese. This issue is related with the teacher's ability to engage in teacher talk, her English proficiency, and her teaching experience. I discuss this issue in relation with other findings at the end of this chapter.

### *Additional Likert Items*

#### *Question 18: The Authenticity of the Language Game*

Question 18 asked the raters if the language game on the screen was authentic.

The raters used the following scale: 1 = *very authentic*; 2 = *to some extent authentic*; 3 = *between '2' and '4'*; 4 = *not very authentic*, and 5 = *not at all authentic*. The mean raw score of the college teacher group (3.58) was higher than those of the student teacher group (2.70) and the junior high school teacher group (2.82). The one-way ANOVA was significant,  $F(2, 101) = 9.30, p = .00$ , partial  $\eta^2 = 0.16$ . Follow-up tests revealed a significant difference in the means between the student teachers and the college teachers ( $p = .00$ ) and also between the college teachers and the junior high school teachers ( $p = .00$ ), but no significant difference between the student teachers and the junior high school teachers ( $p = .81$ ). The college teacher group gave significantly more severe ratings than the other two groups, indicating that they did not think that the language game on the screen was authentic. This issue should be discussed by the three groups so that they can arrive at a more similar understanding of the characteristics that make language games authentic and effective. They especially need to arrive at a common definition of *authenticity*. Two members of the college teacher group pointed out that this

interactional language activity was based on the new sentence pattern and was teacher controlled. Another college teacher said that some of the junior high school students on video were looking at their notebooks and reading the expressions they needed to use while they were interacting with their peers

In the context of the question, I regarded interactional activities as part of a single continuum that links pre-communicative and communicative activities (Littlewood, 1981). In a dialogue-performance, the teacher's control is at a maximum and the learners' creativity is at a minimum. This kind of activity is on one side of the continuum. As learners are given more opportunities to experiment with the language, activities such as contextualized drills and cued dialogues are used. In the more creative types of role-playing, the teacher controls only the situation and the learners' roles, but allows the learners to create the interaction. Improvisation is located on the other side of the continuum. In activities that allow for improvisation, learners are often presented only with a stimulus situation, which they can interpret and exploit in any way they wish. The following is an example (Littlewood, 1981, p. 60):

You are travelling on an underground train (a subway). Suddenly it stops between two stations. At first you take no notice, but soon you all begin to wonder

what is happening. It gets warmer and warmer. You become more and more nervous. After ten minutes, to your relief, the train begins to move again.

One of the examples of improvisation I introduced in my methodology course is: You are participating in a tour in which junior high school students all over the world come to participate. As you get into your bus and find your seat, you meet a student from another country.

In the context of the question for this study, authenticity of an interactional activity is the location of the activity on the continuum.

#### *Question 20: Student Interest in the Language Game*

Question 20 asked the raters if they thought that the game on the screen would be interesting to the junior high school students. The raters used the following scale: 1 = *I strongly disagree*; 2 = *I disagree*; 3 = *between '2' and '4'*; 4 = *I agree*, and; 5 = *I strongly agree*. The mean raw score of the student teacher group (2.16) was lower than those of the college teacher group (2.58) and the junior high school teacher group (2.82). A one-way ANOVA was significant,  $F(2, 101) = 5.21, p = .007$ , partial  $\eta^2 = 0.09$ . There was a significant difference in the means between the student teachers and the junior high school teachers ( $p = .008$ ), but no

significant differences between the college teachers and the junior high school teachers, or between the student teachers and the college teachers. Thus, the student teacher group thought that the game was much less interesting than the junior high school teachers. This would also be a useful area for the three groups of raters to discuss. There might be a measure of validity in the student teachers' judgments if we assume that they understand the junior high school students' views because of their similar ages. On the other hand, an argument can also be made that the junior high school teachers were more familiar with junior high school students and that they are therefore better judges of the students' views.

*Question 21: Appropriacy of the Language Activities and Question 22: Type of Language Activity*

Question 21 asked the raters which type of language activity is most appropriate for junior high school students. The raters were asked to choose one of the following alternatives:

6. Memorized dialogues
7. Contextualized drills in which everything is controlled by the teacher

8. A language game in which students use a new sentence pattern. They can choose any words they like.
9. Cued dialogues
10. Improvisation

Group means were calculated using the raw scores. The junior high teacher group mean ( $M = 3.14$ ) was lower than those of the student teacher group ( $M = 3.46$ ) and the college teacher group ( $M = 3.32$ ).

Twenty-nine (50.88%) of the student teachers selected category 3, while 15 (26.32%) chose category 4. Twelve (63.16%) of the college teachers selected category 3, and 5 (26.32%) selected category 4. Eleven (39.29%) of the junior high school teachers selected category 3, and 10 (35.71%) selected category 4. A chi-square test was conducted to assess whether there were any differences among the ratings of the three groups for each of the six categories. The results of the chi-square test were not significant,  $\chi^2(10, N = 104) = 9.65, p = .47$ .

Most of the student teachers and the college teachers chose category 3, while the junior high school teachers selecting category 3 and category 4 almost the same

number of times. A chi-square test indicated that there were no significant differences between the three groups.

Table 53. *Chi-square Results for Item 21*

Category	Student Teachers	Junior High School	
		Teachers	College Teachers
0	1 (1.75%)	1 (3.57%)	0 (0%)
1	1 (1.75%)	3 (10.71%)	0 (0%)
2	2 (3.51%)	1 (3.57%)	1 (5.26%)
3	29 (50.88%)	11 (39.29%)	12 (63.16%)
4	15 (26.32%)	10 (35.71%)	5 (26.32%)
5	9 (15.79%)	2 (7.14%)	1 (5.26%)

It should be noted that the number of junior high school teachers who chose Category 3 was almost the same as the number of junior high school teachers who chose Category 4. The nature of the control exercised by the teacher is the main criterion for grouping these role-playing activities into the five categories. As the amount of teacher control decreases and becomes less specific, there is increased scope for the learners' creativity. In this respect, the activities can be viewed as part of a single continuum linking pre-communicative and communicative activities (Littlewood, 1981). The category 4 activity is on the borderline between pre-communicative and communicative simulation: The teacher exercises direct control over the meanings that are expressed, but not over the specific language that is used for expressing them. Half of the junior high school teachers thought that the

category 4 activity was suitable for junior high school students. In comparison with the other two groups, more of the participants in the junior high school teacher group were slightly more communicatively oriented than the participants in the other two groups, though, in terms of the mean scores, the student teachers and the college teachers were a little more communicatively oriented, as the junior high school teachers' mean was lower than that of the other two groups. In addition, the standard deviation of the junior high school teachers was larger than that of the other groups. This indicated that the junior high school teachers' ideas concerning appropriate interactive language activities for junior high school students vary a great deal.

Question 22 asked the raters to identify the language activities that they thought the game on the screen represented using the following five categories:

1. Performing memorized dialogues
2. Contextualized drills in which the task is controlled by the teacher
3. A language game in which students practice a new sentence pattern while choosing the vocabulary
4. Cued dialogues
5. Improvisation

Group means were calculated using the raw scores. The college teacher group mean ( $M = 2.95$ ) was lower than those of the student teacher group ( $M = 3.23$ ) and the junior high school teacher group ( $M = 3.00$ ).

Category 3 was selected most frequently by the college teacher, student teacher, and junior high school teacher group [30 (52.6%), 12 (63.2%), 15 (53.6%), respectively], followed by category 4 [20 (35.1%), 4 (21.1%), 8 (28.6%), respectively]. A chi-square test was conducted to assess whether there were any differences among the three groups for the six categories. The results of the test were not significant,  $\chi^2(10, N = 104) = 6.52, p = .77$ .

Table 54. *Chi-square Results for Item 22*

Category	Student Teachers	Junior High School	
		Teachers	College Teachers
0	1 (1.8%)	0 (0%)	0 (0%)
1	1 (1.8%)	3 (10.7%)	2 (10.5%)
2	4 (7.0%)	2 (7.1%)	1 (5.3%)
3	30 (52.6%)	15 (53.6%)	12 (63.2%)
4	20 (35.1%)	8 (28.6%)	4 (21.1%)
5	1 (1.8%)	0 (0%)	0 (0%)

As shown in Table 54, Category 3 was chosen most by the student teacher group (52.6%), the college teacher group (53.6%), and the junior high school teacher group (63.2%). Category 4 was chosen next to the most by every group.

Table 55. Mean Scores of the Three Groups for Items 21 and 22

		Item 21	Item 22
Student teachers	<i>M</i>	3.46	3.23
	<i>SD</i>	0.98	0.82
College teachers	<i>M</i>	3.32	2.95
	<i>SD</i>	0.67	0.85
Junior high school teachers	<i>M</i>	3.14	3.00
	<i>SD</i>	1.21	0.90

The mean scores of the three groups for item 21 are larger than those of the three groups for item 22. This indicates that the language activity shown on the video was more controlled than the raters had expected.

#### Theoretical Consequences of the Results

In this section, I have discussed the results of the *Student Teachers' Video Instruction*. I first identified four independent constructs among the responses from the three rater groups to the 22 questions and the 16 questions in this section of the *Student Teachers' Video Instruction* using the Rasch rating scale model. The four constructs were *Listening and Japanese Use*, *Prereading*, *Using English*, and *Memorization* and the three constructs from the 16 questions were *Explanation* and

*Understanding, Grammar-translation and Habit-formation, and Communicative Practice.*

Most of these questions were paired. For example, one question was “*Is it useful if a teacher introduces a new sentence pattern in English trying to lead students to find out a new rule?*” The paired question was “*The student teacher on the screen tries to introduce a new sentence pattern in English trying to lead students to find out a new rule. Is she effective when her English and her teaching technique are taken into consideration?*”

No statistically significant differences were found among the three rater groups for the four constructs (*Listening and Japanese Use, Prereading, Using English, and Memorization*) that were extracted from the 22 questions (See Figure 18.). Thus, the three groups’ views concerning the ideas related with these 22 questions were not significantly different.

In the discussion of the results of the three constructs (*Explanation and Understanding, Grammar-translation and Habit-formation, and Communicative Practice*) extracted from the 16 questions, a significant difference was found (See Figure 19 and Table 56). This indicated that the three groups’ evaluation of the

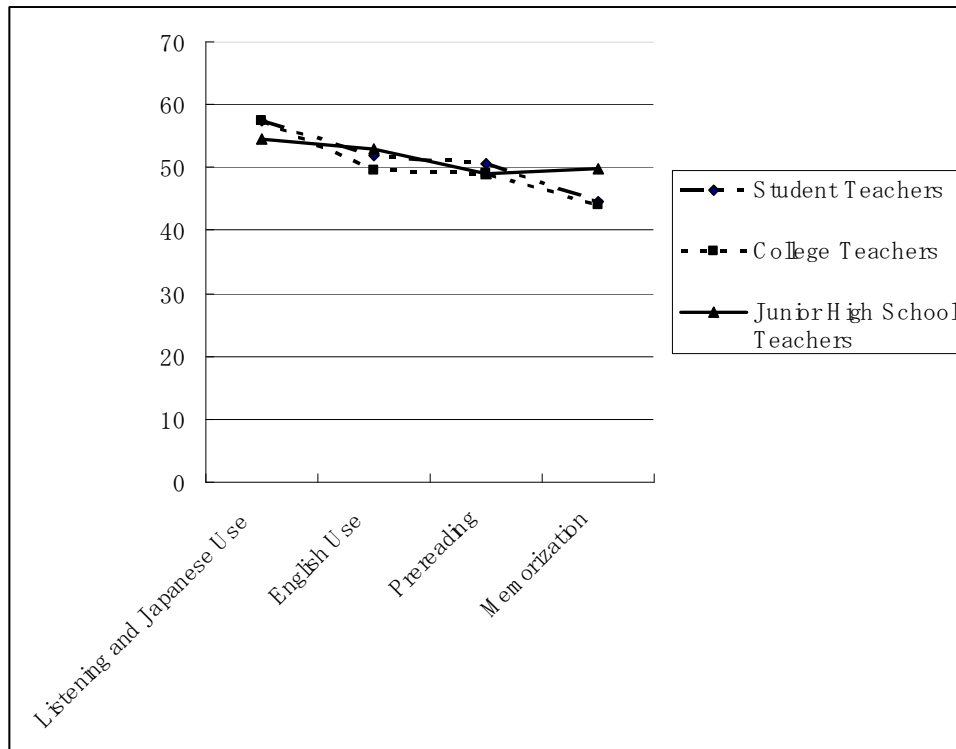


Figure 18. The mean Rasch person ability estimates (CHIPS) of the three usefulness constructs.

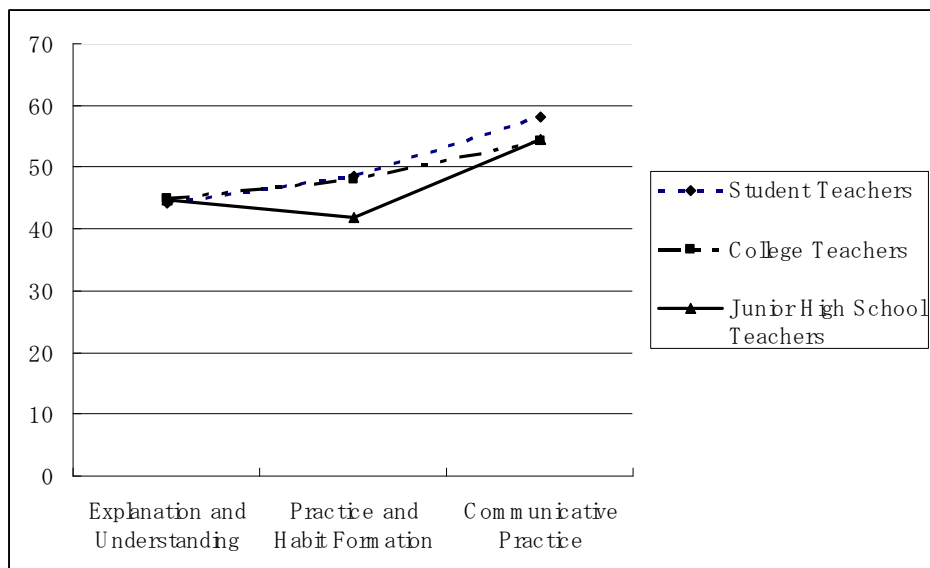


Figure 19. The mean Rasch person ability estimates (CHIPS) of the three effectiveness constructs.

student teachers' teaching differed, though the three groups' views over the ideas on which the student teachers' teaching did not differ.

Table 56. *Statistically Significant Differences Among the Three Groups*

	Explanation and Understanding	Grammar-translation and Habit-formation	Communicative Practice
Student teachers vs. College teachers	Significantly different	Not different	Significantly different
Student teachers vs. Junior high school teachers	Significantly different	Significantly different	Significantly different
College teachers vs. Junior high school teachers	Not different	Significantly different	Not different

The three groups should discuss their differences so that they can share their ideas and the student teachers can learn from the views of more experienced teachers. Also, the questions that were most difficult for the three groups to endorse were points that should be objectives in the training program for student teachers. These questions were Question 27, Question 12, and Question 16 from *Explanation and Understanding*, Question 40 from *Grammar-translation and Habit-formation*, and Question 32, Question 34 from *Communicative Practice* (See Table 57).

The questions (Table 57) should be discussed and the student teachers need to identify weaknesses in their teaching in relation to the above questions. Questions 27, 12, and 16 concern English proficiency as well as the ways the student teachers provided explanations to the junior high school students. Question 40 concerns the student teachers' explanation of the English sentences after they had translated them into Japanese. For example, when a junior high student failed to translate one English sentence into Japanese, the student teacher simply provided a Japanese translation and did not explain the reason(s) why the English sentence was translated in that way, and did not inquire into why the student could not translate the English sentence accurately. Question 32 concerns the qualities of the questions the student teachers asked the junior high school students (i.e., Were the questions pseudo or genuine?) as well as the student teachers' English proficiency. The issues concerned (a) whether the pseudo questions and genuine questions were asked appropriately, (b) whether the questions were too easy for certain students or too difficult for others, and (c) whether the student teachers' English was accurate when they asked questions in English and Japanese.

Table 57. *Teaching Skills in Which the Student Teachers Need Further Training*

Section/Question	Skill Description
<i>Explanation and Understanding</i>	
Question 27	The student teacher introduces the text orally with picture cards before the class starting to read the text. Is she effective?
Question 12	Is the interview game the student teacher on the screen is introducing effective?
Question 16	Is the student teacher's instruction before the game and after the game effective?
<i>Grammar-translation and Habit-formation</i>	
Question 40	The student teacher asked the students to translate the text into Japanese. Is she effective?
<i>Communicative Practice</i>	
Question 32	Overall, is the student teacher effective in asking questions in English or Japanese?

#### Research Question 1: Revised COLT Categories

The first research question asked what categories should be included on the observation checklist that the student teachers, college teachers, and junior high school teachers (student teachers' supervisors) use when observing the student teachers' classroom teaching. The answer to this research question was discussed in

Chapter 4. The COLT was revised based on previous studies I conducted. In this section, I first introduce two viewpoints that informed the revision of the COLT. Next, I discuss how well the Revised COLT worked and the remaining problems with the instrument.

### *The COLT and the Revised COLT*

In this study, the COLT was revised from two points of view. The first point concerned the categories included on the COLT. One of the categories on the original COLT was *Activities and episodes*. When using this category, observers list the teaching activities they observe (e.g., greeting the students, conducting review activities, and introducing a new reading text). On the Revised COLT, however, the observers did not have to write anything in this category because I listed the activities based on my observations of the student teachers' teaching plans, the student teachers' teaching, and experienced teachers' teaching. I also discussed the classroom activities that I observed with the student teachers and junior high school teachers when I visited the junior high schools where the student teachers were participating in the practicum.

The second point concerned the need to make the Revised COLT more user-friendly. This was accomplished by eliminating some of the categories on the Original COLT. When researchers use the COLT, they can use Part A, Part B, or both parts. Part A concerns classroom events at the level of episode and activity, and Part B is focused on the communicative features of verbal exchanges between teachers and students and/or students and their peers as they occur within each episode or activity. The COLT is made up of 73 categories, and most of them represent binary distinctions in instructional practices (e.g., genuine vs. pseudo requests; student-centered vs. teacher-centered participation). Some researchers transcribe interactions between a teacher and students and between peers that occurred in the class. As discussed in Chapter 4, I combined Part A and Part B and also deleted some of the categories that were not observed in the student teachers' or experienced teachers' teaching.

#### *Reliability of the Categories on the Revised COLT*

The participants (the student teachers, the college teachers, the junior high school teachers) used the Revised COLT to check the five student teachers' teaching on the videotape: Revised COLT Block A for student teacher 1, Revised

COLT Block B for student teacher 2, Revised COLT Block C for student teacher 3, Revised COLT Block D for student teacher 4, and Revised COLT Block E for student teacher 5. Cronbach's  $\alpha$  reliability estimates were calculated for the raw data of each block. The reliability (number of categories) for Block A, Block B, Block C, Block D, and Block E were .86 (37), .88 (40), .78 (48), .92 (86), and .83 (53), respectively.

### *The Results of the Pilot Study and Research Question 3*

The participants used the Revised COLT after it had been piloted. One of the main purposes of this study was to obtain feedback from junior high school teachers about the student teachers' performance in the practicum. Two problems were pointed out in relation with this issue. The first problem addressed in this study concerns the lack of clarity in the feedback provided to student teachers, and the second problem is that the junior and senior high school supervisors' feedback regarding the student teachers' teaching performance is generally vague and too brief. In order to solve these problems, the COLT was revised so that it would be more suitable for the three groups of observers who participated in this study (i.e.,

student teachers, college teachers, and junior high school teachers) to use when observing the student teachers' teaching.

For Research Question 3, the ratings that the three groups awarded to the student teachers with the *Revised COLT* while observing them on the video were analyzed and discussed. Although a one-way ANOVA did not reveal any statistically significant differences, four problems with the student teachers' teaching were identified.

The first problem was that when teaching *New sentence pattern*, student teacher 1 mainly explained the pattern and did not interact with the students. The participants evaluated this teacher's performance severely in terms of *SL2 Use*, *Sustained Speech*, *Questions (pseudo)*, and *Purpose (socializing or instruction)*.

The second problem is that student teacher 3, when teaching *Evaluation*, asked fewer questions to the junior high school students than the raters expected, so the raters gave low scores to this teacher. In particular, one junior high school teacher stated that student teacher 3 had fewer interactions and slower interactions than she had expected. The junior high school students had just completed an interactive activity, and the student teacher was expected to encourage the junior high school students to evaluate their performance by asking them whether they

had tried to establish eye contact with the listeners while they were speaking, or if they were able to make themselves understood. If the student teacher had evaluated the students' activity by asking the students in simple English, she might have received higher scores.

The third problem is that in student teacher 4's teaching of *Model/Chorus reading*, the teacher's model reading and the students' practice reading was shorter than the raters had expected, so the teacher's performance was assessed severely. In addition, the student teacher tended to use Japanese rather than English. Thus, this student teacher exhibited problems in terms of *SL2 Use, L2 Use, Questions, and Purpose*.

The fourth problem is that Student Teacher 5, when teaching *Consolidation*, provided few opportunities for the students to speak in Japanese or English. Many junior high school teachers pointed out that the student teachers tended to talk too much; this occurred because they were making an effort to explain everything, even though it would have been preferable to have the junior high school students think more independently and for the student teacher to support their efforts to solve the problem by asking them leading questions.

Because the above four problems corresponded with what the junior high school teachers had pointed out previously, it appears that the Revised COLT allowed for the identification of weaknesses in the teaching of the student teachers in terms of *SL2 Use*, *L2 Use*, *Question*, *Sustained Speech*, and *Purpose*. This finding provides some justification for using the Revised COLT. Although useful, the information recorded on the Revised COLT needed to be supplemented by written and oral comments from several junior high school teachers and college teachers. One strength of the instrument is that it provides observers with a way to focus their discussions by referring to the data they record on the Revised COLT. This process is facilitated by the observers' familiarity with the expressions used on the Revised COLT, which were selected while I observed the student teachers' and some experienced teachers' teaching, and during discussions with student teachers, college teachers, and junior high school teachers.

### *The Role of the Revised COLT*

Audio and/or video recording classroom interactions can be helpful for student teachers as they can observe their own teaching as well as the events taking place in the classroom as a whole. In addition, the use of the video allows all

parties concerned to observe the student teacher's performance multiple times, and this can enable them to notice aspects of the teacher's performance that they were unaware of when watching it in person. This happened, for instance, when the junior high school teachers pointed out that the student teachers tended to pay more attention to academically strong students and active students than to relatively quiet students. The junior high school teachers noticed these quiet students raising their hands to answer questions from the student teachers for the first time when they watched the video. (The video camera located in one of the front of classroom was recording the class). A second example is that the college teachers understood why the student teachers stopped teaching in English and started speaking Japanese when they saw the junior high school students' perplexed expressions on the video. In this case, the information provided by the video recording was necessary because the college teachers usually stood at the back of the class and could therefore not observe the junior high school students' faces.

Even the student teachers, who are not highly trained observers, can use the revised COLT as a classroom observation instrument because it is relatively easy to understand and use. The use of the revised COLT also facilitates discussions about the class they have observed.

By using the language provided by the observation categories or the observation system, the student teachers acquire meta-language that allows them to better comprehend and express their ideas about classroom events, become better able to analyze those events, and thereby more able to identify patterns in the classroom.

#### *Problems with the Revised COLT*

*Categories.* There are two problems with the current categories on the Revised COLT. The first problem concerns the reading subcategory. It was pointed out by one junior high school teacher that silent reading and reading aloud differ in terms of their purposes: students read silently to grasp the meaning of a text and they read aloud to acquire language in the text and to practice pronunciation. For this reason, separate categories should be made for these two types of reading.

The second problem arose when several junior high school teachers commented on the student teachers' English pronunciation. Although there are categories concerning the student teachers' use of English (i.e., how much they speak), there is no place on the COLT to indicate how well or poorly a student teacher is using English. For example, raters give a 1 or 2 when they notice that a

student teacher is using English, but they do not rate how skillfully the student teachers use English. Thus, whether teacher trainers should provide criteria for junior high school teachers to evaluate student teachers' English pronunciation and other subcategories of their English proficiency such as grammatical accuracy is an issue for further study.

### Use of English and Japanese

In this chapter, I have discussed the findings based on the analysis and results of research questions 1-4. After discussing the findings of each research question, there are two issues left to consider. The first concerns the need to increase the student teachers' English proficiency, especially their spoken English proficiency. The second concerns the use of Japanese and/or English in teaching English in Japanese junior high schools.

In junior high school English classrooms, and especially in the beginning stages of learning the language, teachers should speak English as much as possible because they are the primary providers of aural input in the foreign language context. Although linguistic input can also be provided by using CDs, podcasts, videotapes, and DVDs, linguistic input from teachers differs from the input

provided by these other media because teachers can modify the language they provide through repetition, comprehension checks, clarification requests, and confirmations checks. They can also adjust their speaking speed and pausing, offer alternative expressions and synonyms, and provide nonverbal information through their facial expressions, movement, mime, pictures, and realia. This increases the probability that the input will be comprehensible for the learners and thereby minimize the need for translation. This is one of the main reasons why student teachers should be trained to be able to speak English well. With sufficiently high English proficiency, the junior high school English classroom can be a place where students interact with their teacher and peers in English.

Teacher talk is one of the main skills language teachers should develop and student teachers can acquire this skill through teaching experience based on an adequate knowledge of English and knowledge of teaching techniques and methodologies. Based on these ideas, I would suggest that two new programs should be incorporated into teacher licensure programs. The first suggestion is to extend the practicum to as long as one year. Student teachers would stay at a special dormitory where they are only permitted to speak English, and where they would listen to lectures on subjects such as English teaching methodology and

discuss what and how they have taught every day. If teachers can provide junior high school students with comprehensible input, junior high school students would not need to rely on translation or the teacher's use of Japanese, except when the teacher explains English grammar, explains abstract concepts, and/or when reviewing the day's lesson at the end of the class.

## **CHAPTER 8**

### **CONCLUSION**

To the best of my knowledge, this is only the second study in which a version of the COLT was used for teacher training, the first being the study by Block (1992). This study is also the first published report in which Parts A and B were combined. In this chapter, I first present a summary of the findings of the study and this is followed by a discussion of the limitations, and suggestions for future study.

#### **Summary of the Findings**

Four research questions were investigated in this study. The first research question asked what categories should be included on the observation checklist that the student teachers, college teachers, and junior high school teachers (student teachers' practicum supervisors) use when observing the student teachers' classroom teaching. The COLT was revised based on previous studies I had conducted and based on two points of view. The first concerned the categories on the COLT and the second concerned the need to make the Revised COLT more user-friendly. The three groups of raters (the student teacher group, the college

teacher group, and the junior high school teacher group) used the Revised COLT while observing the five student teachers teaching in their practicum on the video. Cronbach  $\alpha$  reliability estimates were calculated for the five sets of ratings using the raw data. The reliability estimates (number of categories) were acceptable: .86 (37), .88 (40), .78 (48), .92 (86), and .83 (53), respectively.

The participants encountered no difficulties using the Revised COLT, and importantly, even the student teachers, who were not highly trained observers and who have little teaching experience, could use the Revised COLT successfully. By using the language provided on the Revised COLT, the student teachers began to acquire meta-language that allows them to better comprehend and express their ideas about classroom events, become more able to analyze those events, and thereby more readily identify patterns in the classroom.

By using the Revised COLT, the three groups of raters identified four specific problems with the student teachers' teaching. The first problem was that when teaching *New sentence pattern*, student teacher 1 mainly explained the pattern and did not interact with the students. The second problem was that student teacher 3, when teaching *Evaluation*, asked fewer questions to the junior high school students than the raters expected. The third problem occurred during student teacher 4's

teaching of *Model/Chorus reading*; the teacher's model reading and the students' practice reading was shorter than the raters had expected. The fourth problem was that Student Teacher 5, when teaching *Consolidation*, provided few opportunities for the students to speak in Japanese or English and the student teacher spoke Japanese more than necessary.

For example, student teacher 1, when teaching *New sentence pattern*, explained the pattern and did not interact with the students. The three groups evaluated this teacher's performance severely in terms of *SL2 Use*, *Sustained Speech*, *Questions (pseudo)*, and *Purpose (socializing or instruction)*. Thus, the Revised COLT helps observers identify specific problems with student teachers' instruction.

There are, however, two problems with the current categories on the Revised COLT. The first problem concerns the reading subcategory. Reading aloud and reading silently to grasp the meaning of a text are different activities with different pedagogical goals, so separate categories should be made for these two types of reading. The other problem concerns whether criteria to evaluate specific aspects of the student teachers' English proficiency (e.g., English pronunciation and grammatical accuracy) should be added to the instrument.

The second research question asked to what degree the student teachers, college teachers, and junior high school teachers' ratings on the Teaching Skills Questionnaire differed. The Teaching Skills Questionnaire was used to determine if there were any differences in the perception of the three groups of raters concerning the 15 teaching skills. Three constructs underlying the questions concerning the 15 teaching skills: *Teaching Mainly in Japanese*, *Teaching Using Easy English*, and *Teaching Using Difficult English*.

Overall, items 11-15 (*Teaching Using Difficult English*), measured skills that were seen by the raters as being the most difficult to do well; thus, teaching skills requiring greater English proficiency are more demanding and more difficult for the student teachers to carry out successfully. Overall, the student teachers perceived each skill as more difficult to teach than the other rater groups. This indicates that the student teachers were differentiated from the older teacher groups because of their limited teaching experience and lower English proficiency. In addition, the student teachers, as young adults, might have found it difficult to manage the younger junior high school adolescents.

The third research question asked to what degree the student teachers, junior high school teachers, and college teachers' ratings of the student teachers' teaching

performances differed when they used the *Revised COLT*. This question was answered by analyzing the three groups' ratings for the *Student Teachers' Teaching on the Video*. No statistically significant differences were found among the three groups for student teachers 1, 2, 3, 4, and 5. Overall, the junior high school teachers gave higher scores to every student teacher except student teacher 3, who received the highest scores from the student teacher group. Especially, the junior high school teachers gave fairly high scores to student teacher 1, who was seen on the video teaching *review* and *new sentence pattern*. The mean scores of the junior high school teachers were high due to the scores they gave student teacher 1's *L2 Use* and *Purpose* (of her L2 use), criteria that they seemed to value, presumably based on their experience teaching in junior high schools.

The fourth research question asked to what degree the student teachers, college teachers, and junior high school teachers' responses differed when they used the *Questions Concerning the Student Teachers' Teaching on the Video*. The scores of the 42 questions, most of which were paired, were analyzed. One group of 22 questions concerned what the student teachers studied in their licensure program, what the college teachers taught in their licensure program, and what the junior high school teachers practiced at school. The four constructs underlying these 22

questions were *Listening and Japanese Use, Prereading, Using English, and Memorization*. There were no statistically significant differences among the groups and this indicates that what the student teachers studied in their licensure program, what the college teachers taught in their licensure program, and what the junior high school teachers practiced at school were similar.

The other group of 16 questions concerned how each group evaluated the five student teachers' teaching. The three constructs underlying these 16 questions were *Explanation and Understanding, Grammar-translation and Habit-formation, and Communicative Practice*. A statistically significant difference was found in the three groups' evaluation of the student teachers' teaching for Questions 27, 12, and 16 from *Explanation and Understanding*, Question 40 from *Grammar-translation and Habit-formation*, and Questions 32 and 34 from *Communicative Practice*.

These questions concern English proficiency, the ways the student teachers provided explanations to the junior high school students (e.g., explanations of the English sentences after they had translated them into Japanese), and the quality of the questions the student teachers asked the junior high school students.

## Limitations

I first discuss the limitations of the whole design, and next consider those limitations in relation with each research question with the instruments used to answer the question.

The first limitation concerns the imbalance in the number of participants in each rater group. The number of participants in the student teacher group, college teacher group, and junior high school teacher group was 57, 19, and 28, respectively. It would have been better if the number of participants in the college teacher and junior high school teacher groups could have been increased to approximately 50 persons per group, as this would have increased the statistical power of the study and produced lower standards errors (i.e., more precise measurement) in their ratings.

The second limitation involved the type and amount of information I collected. I was able to obtain a great deal of information from the results of the statistical analyses, yet there were ideas underlying the participants' ratings that I was likely unaware of. Some of the participants were kind enough to write their ideas about issues related to the study, and whenever I encountered questions, I conducted short, informal interviews with the participants. Providing space on the

answer sheets for the participants to write the reasons for their ratings would have resulted in more interpretable findings and might have shown that similar numerical ratings are sometimes based on different reasons. In short, more information could have been obtained more systematically.

The third limitation concerned the administration of the questionnaires used in this study. Although I was present when the student teachers completed the *Teaching Skill Questionnaire*, the *Revised COLT*, and the *Student Teachers' Videotaped Instruction* and could therefore answer their questions, which was not the case for the college teacher group and the junior high school group. I sent a set of questionnaires with a videotape to each of those participants with an explanation of how to complete each questionnaire and use the Revised COLT. One problem with this approach, however, was that these two groups of participants did not have an opportunity to ask questions while they were completing the questionnaires; thus, some of them might have interpreted some of the questionnaire items differently from what I had intended. If this occurred, it added unwanted error variance to the data.

The fourth limitation involved the quality of the videotaped lessons of the five teachers' teaching while they were teaching in their practicum at public junior high

schools. For financial reasons, I bought a relatively inexpensive video camera, and although I equipped the camera with a microphone, the sound quality was not particularly good, so the voices of the junior high school students on the video were difficult to understand at times. This technical limitation might have interfered with the raters' attempts to clearly understand the interactions that took place on the videotapes.

The final limitation concerned the procedures followed by the participants when answering the 42 questions on the *Student Teachers' Videotaped Instruction*. One set of 22 questions concerned what the student teachers studied in their licensure program, what the college teachers taught in their licensure program, and what the junior high school teachers practiced at school. The participants answered these questions after watching the five student teachers' teaching on the video and evaluating their performances with the Revised COLT. Their answers to the 22 questions might have been influenced by their observation of the five teachers' teaching; thus, asking them to answer the questions before watching the video might have resulted in more unbiased responses. One example is: There are 42 questions on the *Questions Concerning the Student Teachers' Teaching on the Video*, 22 of which questions asked the participants about usefulness of teaching

methods and techniques. One of the constructs was *Listening and Japanese Use* underlying question 5 (*In these two scenes, what the students did was just listening except on some occasions. Is this useful for review?*), question 7 (*Is it useful for students to translate one Japanese sentence into English using the new sentence pattern?*), question 13 (*Is it useful to ask questions in Japanese about the text?*), 30 (*Is it useful to let students translate the text into Japanese?*), and question 39 (*Is it useful to explain about English grammar in Japanese?*). The item difficulty hierarchy was, from the easiest to the most difficult item, Question 7, 13, 39, and 5. What made Question 5 different from the other questions was the use of English. Most of the raters chose *Not useful* or *Not very useful* for question 5. The reason for this severe rating might be that the raters were not satisfied with the student teachers' English, which was clearly weak, even though they were supposed to rate the usefulness of the teaching methods.

### Suggestions for Future Research

The COLT is a useful instrument for analyzing student teachers' teaching because it can provide university/college teachers and junior high school teachers, and student teachers with specific records of their perceptions, which can be used

immediately or at a later time to discuss student teachers' teaching in specific ways. Thus, the first suggestion is that teachers and researchers working in this area should further revise and develop the COLT so that it better suits the needs of all parties taking part in the teaching practicum. This might, for instance, take the form of in-depth validation studies.

Second, if different groups, such as student teachers, college teachers, and junior high school teachers, are to be compared using statistical means, the number of participants' in each group should be both larger than those in this study, and relatively equal. This would produce somewhat more generalizable and precise findings. In addition, adding more participants would allow for the investigation of more independent variables using criteria such as English proficiency (high and low proficiency groups), or the amount of micro-teaching experience the teachers have had prior to the practicum (groups with a little or a great deal of experience).

Third, future researchers should consider using a mixed-methods design incorporating both quantitative and qualitative approaches, as this would allow them to arrive at more well balanced findings and obtain greater insight into the results of any statistical analyses.

Fourth, researchers should consider focusing on the use of Japanese and English on the Revised COLT. This might involve determining when student teachers use Japanese and their purposes in using the language as well as the student teachers' use of English and their purposes. Because these checks are still at the categorical level, and therefore yield limited information about the raters' opinions, discussions based on what observers check with the Revised COLT, are necessary. Investigations could also be focused on whether it is preferable to use English where Japanese is used and whether it is preferable to use Japanese where English is used. Another area for investigation concerns the problems student teachers encounter with English in the classroom. This could result in an English for Specific Purposes syllabus designed for use with student teachers before they take part in the practicum.

The final suggestion concerns the teaching of English in Japanese elementary schools. Developing and validating a simplified version of the COLT would be useful in facilitating elementary school teachers' ability to carry out English teaching activities. Because this is task for which neither their training or experience has prepared them for, gaining feedback on their performance and discussing that performance with others is needed. In order to adopt the Japanese

COLT or the original COLT, we need to delete, add, and change some categories.

For example, under the present Course of Study, Japanese public elementary school children do not read and write in English, so the subcategories of *Student modality* need to be deleted. In order to delete, add, and change categories appropriately, a long period of observation of various classes at elementary school, interviews with elementary school teachers and discussion are necessary.

Elementary school teachers are currently using feedback sheets in which children write down what they think about the English activity class they have attended. The children are asked to respond to specific issues when completing this task. For example, they are asked to consider whether they have tried to communicate with their peers. Such issues are drawn primarily from the Course of Study for 'Foreign Language Activities at Public Elementary School', but some also come from teaching policies that apply to this course as well as other courses such as Japanese and arithmetic. Through careful observation, interviews, and discussion, we might identify criteria to delete, change, and add in order to produce a new version of the COLT that elementary school teachers can use to analyze their teaching.

Being faithful to the Course of Study is fine, but it is not enough if the goal is to reform English education at the elementary school level. As one of the purposes of 'Foreign Language Activities at Public Elementary School' is to cultivate children's attitude to learn English and other foreign languages and to take an interest in foreign cultures, categories allowing observers to check children's attitudes must be carefully added to the Japanese COLT. The precise categories to add will be a controversial issue to solve.

### Final Conclusions

In order to improve teacher licensure programs provided by universities and colleges, it is indispensable for student teachers, university and college teachers involved in teacher education programs, and junior high school teachers who supervise student teachers to discuss student teachers' teaching. As the results of this study have shown, using the Revised COLT prompts them to discuss specific problems with the student teachers, and these discussions can be fruitful in a number of ways. For instance, Student Teacher 3, when teaching *Evaluation*, asked fewer questions to the junior high school students than the raters expected and Student Teacher 5, when teaching *Consolidation*, provided few opportunities for

the students to speak in Japanese or English and the student teacher spoke Japanese more than necessary. Thus, practicum supervisors could provide Student 3 with the following feedback: “You need to interact with the students and let them speak more.” This kind of feedback is useful, but still vague. With the Revised COLT, more specific feedback can be given to the student teachers.

The practicality of the Japanese COLT or a new version of the instrument is an important issue; it is necessary to simplify the Japanese COLT further if it is to be used as the basis of the one-hour discussion between the student teacher, college teacher, and junior high school teacher just after the student teacher’s teaching. The degree of simplification depends on the length and focus of the discussion. If each participant can watch the lesson recorded on videotape or in person, they do not have to simplify it, though they may have to change categories based on the lesson.

At present, student teachers in Japan participate in a three- to four-week practicum, which is too far short for adequate teacher training. An alternative is to offer teacher training to new graduates in the first year after they obtain a full-time teaching position. This system is currently being carried out in Japan. New teachers go to junior high school or senior high school to teach and they attend training sessions. However, one problem with this system is that only those who have

passed paper tests and interviews and have obtained full-time teaching positions can attend. This is not a part of the teacher licensure program, but part of high school teachers' job duties.

Student teachers registered in the teacher licensure program should have a longer practicum. However, there are impediments to this suggestion. The first concerns how many student teachers high schools can accommodate every year. Under the present system, lengthening the practicum seems difficult. However, student teachers registered in the teacher licensure program could be an assistant teacher in the Foreign Language Activities at Public Elementary School course. They will be helpful, as they have more knowledge of the English language than elementary school teachers generally, and will help them acquire the ability to conduct classes in English. Much of the English used in elementary schools is concrete, so communicating in English with the aid of gestures, picture cards, realia, and facial expressions is possible for student teachers.

Student teachers need specific feedback and advice from their practicum supervisors. They also need to discuss that feedback with their university or college supervisors and hear their advice about how to improve their teaching. In this regard, I hope that more frequent exchanges will be made between university and

college teachers, and junior high school teachers and that the Revised COLT will play a role in making these exchanges worthwhile for all parties. The use of the COLT for observing peers' teaching and his/her own teaching should also be adopted so that novice teachers can provide feedback to their peers and receive feedback from them as well as from their supervisors in a systematic way.

## REFERENCES

- Allen, P. (1983). A three-level curriculum model for second language education. *Canadian Modern Language Review*, 40, 23-43.
- Allen, P., Bialystok, E., Cummins, J., & Mougeon, R. (1982). *The development of bilingual proficiency*. ( An interim report on the first year of research)). Toronto, Ontario: Modern Language Centre, Ontario Institute for Studies in Education, University of Toronto.
- Allen, P., Fröhlich, M., & Spada, N. (1984). The communicative orientation of language teaching: An observation scheme. In J. Handscombe, R. S. Orem, & B. Taylor (Eds.), *On TESOL: '83: The question of control* (pp. 231-252). Alexandria, VA: TESOL.
- Allison, D. (1998). Investigating learners' course diaries as explorations of language. *Language Teaching Research*, 2(4), 24-47.
- Allwright, R. L. (1980). Turns, topics and tasks: patterns of participation in language teaching and learning. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language acquisition research* (pp. 165-187). Rowley, MA: Newbury House.
- Allwright, D. (1988). *Observation in the language classroom*. London: Longman.

Allwright, R. L., & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Anonymous. (1993). *Kyoikuyosei-katei-ni okeru kyoiku-jisshu-to-kyoikugenba-ni okeru shoninsha-kenshu-tou-ono naiyokanren-wo fumaeta kaizen-karikyuramu-no kaihatsu [ Revised curriculum concerning a practicum in a teacher licensure program and training for beginning teachers ]*. Toyama: The Faculty of Education of Toyama University.

Atkinson, D. (1987). The mother tongue in the classroom: a neglected resource? *ELT Journal*, 41(4), 241 - 247.

Bachman, L., & Palmer, A. (1981). A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading. In A. Palmer (Ed.), *The construct validity of tests of communicative competence* (pp. 149-165). Alexandria, VA: TESOL.

Bailey, K. M., & Ochsner, R. (1983). A methodological review of the diary studies: windmill tilting or social science? In K. M. Bailey, M. H. Long, & S. Peck (Eds.), *Second language acquisition studies* (pp. 188-198). Rowley, MA: Newbury House.

- Bailey, K. M., & Ochsner, R. (1983). *A methodological review of the diary studies: windmill tilting or social science?* Rowley, MA: Newbury House.
- Bailey, K. M. (2001). Action research, teacher research, and classroom research in language teaching. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 489-498). Boston: Heinle and Heinle.
- Banerjee, J. V. (1996). UCLES Report: The design of the classroom observation instruments, unpublished report commissioned by the University of Cambridge Local Examinations Syndicate (UCLES). Cambridge: UCLES.
- Barnes, D., Britton, J., & Rosen, H. (1969). *Language, the learner and the school*. Harmondsworth: Penguin.
- Bejarano, Y. (1987). A cooperative small-group methodology in the language classroom. *TESOL Quarterly*, 21(3), 485-504.
- Bellack, A. A., Kliebard, H. M., Hyman, R. T., & Smith, F. L. (1966). *The language of the classroom*. New York: Teachers College Press.
- Bellingham, L. (2004). In there language acquisition after 40? Older learners speak up. In P. Benson & D. Nunan (Eds.), *Learners' stories* (pp. 4-21). Cambridge: Cambridge University Press.
- Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20(3), 463-494.

- Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26(2), 319-343.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste* (R. Nice, Trans.). London: Sage.
- Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Cambridge, MA: Harvard University Press.
- Breen, M., & Candlin, C. (1980). The essentials of a communicative curriculum in language teaching. *Applied Linguistics*, 1, 1-47.
- Brown, C. (1985). *Two windows on the classroom world: Diary studies and participant observation differences*. Washington, D.C.: TESOL.
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.
- Burns, A. (1998). *Collaborative action research for English language teachers*. Cambridge: Cambridge University Press.

- Campbell, C. C. (1996). Socializing with the teachers and prior language learning experience: A diary study. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education* (pp. 201-223). New York: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases for communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.
- Carrasco, R. L. (1981.). *Expanded awareness of student performance: A case study in applied ethonographic monitoring in a bilingual classroom*. Rowley, MA: Newbury House.
- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly, 21*, 461-481.
- Cathcart, R. (1986). *Situational differences and the sampling of young L2 children's school language*. Rowley, MA: Newbury House.
- Cathcart, R., Strong, M. A., & Wong-Fillmore, L. (1979). *The social and linguistic behavior of good language learners*. Washington, D.C.: TESOL.
- Cazden, C. B. (1986). *Classroom discourse*. New York: Macmillan.
- Cazden, C. B., John, V. P., & Hymes, D. (Eds.). (1972). *Functions of language in the classroom*. New York: Teachers College Press.

- Chaudron, C. (1988). *Second language classrooms*. Cambridge: Cambridge University Press.
- Chaudron, C. (1980). Review article: Those dear old golden rule days... *Journal of Pragmatics*, 4, 157-172.
- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27, 29-46.
- Cleghorn, A., & Genesee, F. (1984). Languages in contact: An ethnographic study of interaction in an immersion school. *TESOL Quarterly*, 18(4), 595-625.
- Cohen, L., & Manion, L. (1985). *Research methods in education*. London: Croom Helm.
- Collins, J. (1993). *Determination and contradiction: An appreciation and critique of the work of Pierre Bourdieu on language and education*. Cambridge: Polity Press.
- Cotterall, S. (2004). 'It's just rules...that's all it is at this stage...' In P. Benson & D. Nunan (Eds.), *Learners' stories* (pp. 101-118). Cambridge: Cambridge University Press.
- Day, R. R. (1990). Teacher observation in second language teacher education. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 43-61). Cambridge: Cambridge University Press.

- Donate, R., & Adair-Hauck, B. (1992). Discourse Perspectives on Formal Instruction. *Language Awareness, 1*(2), 73-89.
- Doughty, C. J., & Varela, E. (1998). Communicative focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.
- Duff, P. A. (1996). Different languages, different practices: Socialization of discourse competence in dual-language school classrooms in Hungary. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education* (pp. 407-443). New York: Cambridge University Press.
- Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.
- Ellis, R. (1994). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Enright, D. S. (1984). *The organization of interaction in elementary classrooms*. Washington, D.C.: TESOL.
- Erickson, F. (1977). Some approaches to inquiry in school-community ethnography. *Anthropology and Education Quarterly, 8*(2), 58-69.
- Fanselow, J. F. (1977). Beyond Rashomon-conceptualizing and describing the teaching act. *TESOL Quarterly, 11*, 17-39.

- Farrar, M. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, 28(1), 90-98.
- Fazio, L., & Lyster, R. (1998). Immersion and submersion classrooms: A comparison of instructional practices in language arts. *Journal of Multilingual and Multicultural Development*, 19(4), 303-317.
- Field, A (2005). *Discovering statistics using SPSS*. London: Sage.
- Flanders, N. A. (1970). *Analyzing teaching behavior*. Reading, MA: Addison-Wesley.
- Fröhlich, M., Spada, N., & Allen, P. (1985). Differences in the communicative orientation of L2 classrooms. *TESOL Quarterly*, 19, 27-57.
- Fujimori, C. (1995). Riidingu-no-tame-no goi jisho shidou [ Teaching vocabulary and how to use a dictionary for reading ]. *The English Teachers' Magazine*, 43(11), 17-19.
- Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Erlbaum.
- Gass, S. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224-255). Oxford: Blackwell.

- Gebhard, J. G., Gaitan, S., & Oprandy, R. (1990). Beyond prescription: The student teacher as investigator. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 16-25). Cambridge: Cambridge University Press.
- Green, S. B., & Salkind, N. J. (2008). *Using SPSS for Windows and Macintosh*. Upper Saddle River, NJ: Prentice Hall.
- Grittner, F. M. (1968). Letter to the editor. *Newsletter of the National Association of Language Laboratory Directors*, 3, 7.
- Han, Z. (2002). A study of impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.
- Hatch, E. (1978). Discourse analysis and second language acquisition. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 401-435). Rowley, MA: Newbury House.
- Hidaka, S. (1989). Yori jissenteki-na eigokakyouikuhou-eno shian [ What junior college students can prepare in their teaching method class for their practicum? ]. *Bulletin of Hokuriku Gakuin Junior College*, 21, 151-181.
- Hilleson, M. (1996). "I want to talk to them but I don't want them to hear": An introspective study of second language anxiety in an English-medium school. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education* (pp. 248-275). New York: Cambridge University Press.

- Hymes, D. H. (1970). On communicative competence. In J. Gumperz & D. H. Hymes (Eds.), *Directions in sociolinguistics* (pp. 35-71). New York: Holt, Rinehart and Winston.
- Ibe, T. (1993). Gakkoudankaibetsu-ni-mita-eigokyouiku-no-gaikan [ The survey over people's opinions concerning English education according to learners' age ]. *Eigokyoiku*, 42, 9-12.
- Isoyama, S. (1995). Jugyou-deno nihongo-no-shiyou-ha zaiaku-ka [Is it wrong to use Japanese in the English class?]. *The New English Classroom*, No. 308, 19-24.
- Jimbo, H., Hirono, T., Yamazaki, A., Ishida, M., Hisamura, K., & Sakai, S. (2004). *Developing English teacher competency: An empirical study of pre-service teachers, training and curriculum*. Tokyo: JACET SIG on English Education.
- Johnson, K. (1982). *Communicative syllabus design and methodology*. Oxford: Pergamon Press.
- Kaneda, M. (1986). *Eigo-no jugyo-bunseki [English classroom research in Japan]*. Tokyo: Taishukan.
- Kern, R. (1994). The role of mental translation in second language reading *Studies in Second Language Acquisition*, 16(4), 441-460.

- Kizuka, M. (1996). The importance of action research in pre-service education. *Bulletin of Shikoku-eigo-kyoiku-gakkai*, 16, 11-20.
- Kobayashi, T. (1998). How to improve pre-service teacher training. *Gakuen*, 697, 21-29.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Kumaravadivelu, B. (1999). Critical classroom discourse analysis. *TESOL Quarterly*, 33(3), 453-484.
- Kurosaki, T. (2002). 'Kyoiku-jisshu-no hyouka'ni-kansuru kenkyu [ A study of evaluation of student teachers ]. *Bulltin of the Faculty of Education of Okayama University*, 121, 143-150.
- Larsen-Freeman, D. (1996). The changing nature of second language classroom research. In J. Schachter & S. Gass (Eds.), *Second language classroom research: Issues and opportunities* (pp. 157-170). Mahwah, NJ: Erlbaum.
- Leinhardt, G., & Smit, D. (1984). *Expertise in mathematics instruction: Subject matter knowledge* (Paper presented at the annual meeting of the American educational research association). New Orleans.
- Leung, C.-Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language*, 14(1), 66-81.

- Lin, A. M. Y. (1999). Doing-English-lessons in the reproduction or transformation of social worlds? *TESOL Quarterly*, 33(3), 393-412.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3(2), 103-122.
- Linacre, J. M. (2009). Winsteps (Version 3.68.0) [Computer Software]. Chicago: Winsteps.com.
- Lindblad, T. (1969). *Implicit and explicit-an experiment in applied psycholinguistics*. GUME-Pgojektet-1. Gothenburg: Pedagogiska Institutionen, Lärarhögskolan: Göteborg, Göteborgs Universitet, Engelska Institutionen.
- Littlewood, W. T. (1981). *Communicative language teaching: An introduction*. Cambridge: Cambridge University Press.
- Long, M. H., Adams, L., McLean, M., & Castaños, F. (1976). Doing things with words-verbal interaction in lockstep and small group classroom situations. In J. F. Fanselow & R. Crymes (Eds.), *On TESOL '76* (pp. 137-153). Washington, D.C.: TESOL.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Richie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Amsterdam: John Benjamins.

- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37-66.
- Makita, M. (1996). Ima-da-kara chugakusei-ni yomasetai kyouzai [Best reading materials for junior high school students at the best time]. *Modern English Teaching, 33*(4), 12-14.
- Malcolm, D. (2004). An Arabic-speaking English learners' path to autonomy through reading. In P. Benson & D. Nunan (Eds.), *Learners' stories* (pp. 69-82). Cambridge: Cambridge University Press.
- Markee, N. (1996). Making second language classroom research work. In N. J. Schachter & S. Gass (Eds.), *Second language classroom research: Issues and opportunities* (pp. 117-155). Mahwah, NJ: Erlbaum.
- McPherson, P. (1997). Action research: Exploring learner diversity. *Prospect: A Journal of Australian TESOL, 12*(1), 50-62.
- Ministry of Education. (1989). *Course of study for junior high schools: Foreign languages—English*. Tokyo: Tokyoshoseki.
- Ministry of Education. (1990). *Course of study for senior high schools: Foreign languages—English*. Tokyo: Tokyoshoseki.

- Mitsuo, S. (2000). Evaluation of student teachers' teaching and feedback from practicum supervisors at junior high school part I. *Bulletin of Tokyo Junshin Women's College*, 4, 55-65.
- Mitsuo, S., & Uchida, I. (1998). What aspects of English should student-teachers from junior colleges improve before their practicum? *Bulletin of Nihon University of College Pharmacy*, 36, 8-31.
- Miyasako, N. (2002). Koukousei-no ondoku-to-eigoryoku-ha kankei-ga aruka [What effects does oral reading have on English language ability and metacognition of oral reading for Japanese senior high school students?]. *STEP Bulletin*, 14, 14-25.
- Mohatt, G. V., & Erickson, F. (1981). *Cultural differences in teaching styles in an Odawa school: A sociolinguistic approach*. Rowley, MA: Newbury House.
- Mori, C. (2003). Eigogakushu-no-naka-no ondoku-no-ichiduke [Roles of reading aloud in acquiring English]. *Eigokyoiku [The English Teachers' Magazine]*, 52(6), 8-9.
- Moskowitz, G. (1971). Interaction analysis—A new modern language for supervisors. *Foreign Language Annals*, 5, 211-221.
- Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

- Negishi, M. (1995). Riidingu-no tesuto-to hyouka [Test and evaluation for reading]. *The English Teachers' Magazine*, 43(11), 29-31.
- Newmark, L., & Riebel, D. A. (1968). Necessity and sufficiency in language learning. *Internatonal Review of Applied Linguistics*, 6, 145-164.
- Nunan, D. (1990). *Action research in the classroom*. Cambridge: Cambridge University Press.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Nukui, M., Miura, K., & Yoshida, M. (2002). A trial of portfolio assessment to student teaching. *Kyoiku-jissenkenkyu of Chiba University*, 9, 197-208.
- Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *The Modern Language Journal*, 87(iv), 519-533.
- Omaggio, A. C. (1979). Pictures and second language comprehension: Do they help? *Foreign Language Annals*, 12, 107-116.
- O'Sullivan, B., Wier, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing*, 19(1), 33-56.
- Palmer, H. E. (1921). *The oral method of teaching language*. Cambridge: W. Heffer & Sons.

- Perfetti, C. (1985). *Reading ability*. Oxford: Oxford University Press.
- Pica, T. (1987). The selective impact of classroom instruction on second-language acquisition. *Applied Linguistics*, 6, 214-222.
- Politzer, R. L. (1970). Some reflections on "good" and "bad" language teaching behaviors. *Language Learning*, 20, 31-43.
- Prabhu, N. S. (1979) The teaching of English and notions about communication. In J. C. Richards (Ed.), *Applications of linguistics to language teaching, RELC Anthology 6*. Singapore: Singapore University Press.
- Richards, J. C., & Schmidt, R. W. (1983 ). Conversational analysis. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 117-154). London: Longman.
- Rossiter, M. J. (2003). The effects of affective strategy training in the ESL classroom. *TESL-EJ*, 7(2), 1-20.
- Sano, M. (1995). Riidingujugyou-ni okeru oraru-intaakushon [Oral interaction in reading class]. *The English Teachers' Magazine*, 43(11), 8-10.
- Sataporn, S., & Lamb, M. (2004). Accommodation zone. In P. Benson & D. Nunan (Eds.), *Learners' stories* (pp. 119-133). Cambridge: Cambridge University Press.

- Savignon, S. (1972). *Communicative competence: An experiment in foreign-language teaching*. Philadelphia: Center for Curriculum Development.
- Saville, K. (2002). Online discussion techniques for teacher trainers and language learners. In P. Barker & S. Rebelsky (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002* (pp. 1748-1749). Chesapeake, VA: AACE.
- Saville-Troike, M. (1982). *The ethnography of communication: An introduction*. Baltimore: University Park.
- Scherer, G. A. C., & Werthimer, M. (1964). *A psycholinguistic experiment in foreign language teaching*. New York: McGraw Hill.
- Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.
- Schumann, F. M. (1980). *Diary of a language learner: A further analysis*. Rowley, MA: Newbury House.
- Siegel, S., & Castellan, J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

- Shaw, P. A. (1983). *The language of engineering professors: A discourse and registral analysis of a speech event*. Unpublished Ph. D. dissertation, University of Southern California, Los Angeles.
- Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse*. London: Oxford University Press.
- Smith, P. D. (1970). *A comparison of the cognitive and audiolingual approaches to foreign language instruction: The Pennsylvania language project*. Philadelphia: Center for Curriculum Development.
- Snow, M. A., Hyland, J., Kamhi-Stein, L., & Yu, J. H. (1996). U.S. language minority students: Voices from the junior high classroom. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education* (pp. 304-317). New York: Cambridge University Press.
- So, S., & Dominguez, R. (2004). Emotion process in second language acquisition. In P. Benson & D. Nunan (Eds.), *Learners' stories* (pp. 42-55). Cambridge: Cambridge University Press.
- Spada, N., & Fröhlich, M. (1995). *Communicative orientation of language teaching observation scheme*. Sydney: National Centre for English Language Teaching and Research (NCELTR), Macquarie University.
- Spada, N., & Lyster, R. (1997). Microscopic and macroscopic views of second language classrooms. *TESOL Quarterly*, 31, 787-795.

- Stern, H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in Honor of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Takaki, N. (1997). A study of EFL practice teaching as seen in mutual observations of student-teachers and supervisors. *Bulletin of the Faculty of Education of Kumamoto University*, 46, 231-243.
- Tsui, A. B.-M. (1985). Analyzing input and interaction in second language classrooms. *RELC Journal*, 16, 8-32.
- Tsui, A. B. M. (1996). Reticence and anxiety in second language learning. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education* (pp. 145-167). New York: Cambridge University Press.
- Ulichny, P. (1996). Performed conversations in an ESL classroom. *TESOL Quarterly*, 30(4), 739-764.

- Umino, T. (2004). Learning a second language with broadcast materials at home: Japanese students' long-term experiences. In P. Benson & D. Nunan (Eds.), *Learners' Stories I* (pp. 134-149). Cambridge: Cambridge University Press.
- van Lier, L. (1982). *Analyzing interaction in second-language classrooms*. Unpublished Ph. D. Dissertation, University of Lancaster, Lancaster, England.
- van Lier, L. (1988). *The classroom and the language learner: Ethnography and second language classroom research*. London: Longman.
- van Lier, L. (1996). Conflicting voices: Language, classrooms, and bilingual education in Puno. In K. M. Bailey & D. Nunan (Ed.), *Voices from the language classroom: Qualitative research on language education* (pp. 363-387). New York: Cambridge University Press.
- Wallace, C. (1992). *Reading*. Oxford: Oxford University Press.
- Wallace, M. J. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.
- Watson-Gegeo, K. A. (1988). Ethnography in ESL: Defining the essentials. *TESOL Quarterly*, 22(4), 575-592.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

- Wilkins, D. A. (1976). *Notional syllabuses*. Oxford: Oxford University Press.
- Wilkinson, L. C. (Ed.). (1982). *Communicating in the classroom*. New York: Academic Press.
- Wilson, S. (1977). the use of ethnographic techniques in educational research. *Review of Educational Research*, 47(1), 245-265.
- Wong-Fillmore, L. (1980). Learning a second language: Chinese children in the American classroom. In J. E. Alatis (Ed.), *Current issues in bilingual education. Georgetown University Round Table on Languages and Linguistics 1980* (pp. 309-325). Washington, D.C.: Georgetown University Press.
- Wong-Fillmore, L. (1985). *When does teacher talk work as input?* Rowley, MA: Newbury House.
- Wright B. D., & Masters G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Yalden, J. (1983). *The communicative syllabus: Evolution, design and implementation*. Oxford: Pergamon Press.

## **APPENDICES**





**APPENDIX B**  
**THE STUDENT TEACHERS' PRACTICUM**  
**TEACHING QUESTIONNAIRE**

1. How many periods did you teach? What level did you teach?
  2. Which textbook did you use and which section did you teach? What sentence structures did you teach?
  3. Answer about the structures you taught. Was there any part of English grammar that was difficult to teach? If there was any, please describe it in detail and also try to write reasons for the difficulty you found.
  4. How did you teach these structures that you have mentioned in Question 3?
  5. When did you think that you need to improve your English: while you were teaching or preparing for the next class? "When I was thinking of good examples for the new structure" is an example.
  6. Did the Japanese teacher of English give any advice about your English or your teaching?
- 7/8. The following items were asked in questions 7 and 8:
- 1 Pronounce very well at the word level
  - 2 Pronounce very well at the sentence level
  - 3 Instruct students in English
  - 4 Provide good examples of the new structure
  - 5 Explain new structures well
  - 6 Answer questions about the Japanese meaning of an English word
  - 7 Answer questions concerning grammar
  - 8 Translate English texts into Japanese if necessary
  - 9 Make an oral introduction of a new sentence pattern
  - 10 Do a good model reading

- 11 Ask questions in English about the text and have students answer in English
- 12 Introduce context in your own words in English
- 13 Notice students' serious English mistakes and correct them immediately
- 14 Plan an authentic and interesting language activity

## APPENDIX C

### STUDENT TEACHERS' ENGLISH PROFICIENCY

The problems that the participants pointed out in answering questions 2, 3, and 5 on the questionnaire (See Appendix A) are categorized into four groups: pronunciation, vocabulary, grammar, and other aspects. The problems are listed below.

#### Pronunciation

1. Some participants could not do the model reading well. They are generally not good at pronunciation at the sentence level.
2. While instructing junior high school students in English and/or having students repeat new words after them, some participants thought that their own pronunciation was not good enough to serve as a model. For instance, they pronounced some English words by adding a vowel at the end of a closed syllable and inserting a vowel between each consonant in a consonant cluster.
3. Some participants did not understand some basic rules of English pronunciation.
4. Some participants found themselves pronouncing English very inaccurately. They had to be consciously aware of the movements of their lips and tongues. They wished they had practiced English pronunciation much more before the practicum.

#### Vocabulary

1. Some of the participants could not provide appropriate English words for Japanese ones when asked by the junior high school students.
2. Some of the participants could not explain the usage of basic verbs like *take* and *bring*.
3. Some of the participants could not identify some of the junior high school students' misspellings written in their notebooks or on the blackboard.

The three problems concerned with pronunciation were also identified through observing the videotapes. The student teachers' English pronunciation was sometimes unnatural, unclear, and not good enough to be understood by one of the American college teachers. As for item 4, some participants looked very uncomfortable when they were reading or speaking in English. Their facial expressions might not have given a good impression to the junior high school students, though their attitude towards teaching was sincere and they appeared dedicated on the video. Some student teachers' techniques related with pronunciation should be included in a modified list.

The three problems related to vocabulary were observed on the video. Problems concerning items 5 and 6 on the questionnaire also occurred in the other classes when their teaching was not videotaped.

## Grammar

1. Lack of authentic contexts in which example sentences are given: "Is that A or B?" "What do you . . . ?," "I must/ mustn't . . .,"
2. Tag questions
3. Poor understanding of structures and grammar: What to infinitive; *I don't know where to go.*
4. Present perfect verb tense
3. Structures and grammar that student teachers found difficult to explain: The usage of the referential "it": "This is a nice computer. It is made by IBM."
4. Lack of authentic language activities in which target sentences are included: "There is/are..."

One of the most difficult problems the participants had concerned teaching grammar. The participants' answers to question 3 on the questionnaire was translated and the answers were categorized into four groups (See Table 4). Most of the participants were familiar with the structures. On the videotape, they were easily able to translate English sentences with these structures into Japanese if they found it necessary to do so, and they could answer the grammatical exercises that were focused on these structures. It was much harder for them to explain these structures and prepare authentic situations in which these structures were used. Some of the participants' introductions of new sentence patterns were still at the level of recognition and they did not provide authentic situations with new sentence patterns or explanations of its usage.

One possible reason for their difficulties in teaching grammar is that in any language class at senior high school and college, it is often taken for granted that the students understand these basic structures because they can translate English sentences using these structures; however, the proper use of such structures are not frequently taught except in some conversational classes and writing classes. The participants had not yet understood these structures completely in spite of the explanations given in their secondary school English classes. Their knowledge of grammar may not have become part of their implicit knowledge (Ellis, 1994). The circumstances under which the participants obtained their knowledge of grammar and their learning style were not encouraging them to acquire implicit knowledge. Although they were in formal and academic settings when they learned English grammar, they could have tried to use these forms communicatively by interacting with classmates if new structures were introduced in communicative tasks (e.g., information gap tasks). The participants were not in the habit of learning new structures in terms of their usage in authentic situations. When new structures were introduced in their high school lessons, the focus was strictly on accuracy and on Japanese meanings of sentences using these new structures.

## Other Problems

1. When a junior high school student asked some student teachers to give an authentic example of how to use an idiom in a sentence which the participants had just introduced, they could not give one.
2. Some participants could not make even simple comments on junior high school students' performance in English, though they knew that it was important to praise junior high school students when they did their best.

The following issues were raised by six of the participants. I was able to recognize the same problems on the video.

Every student teacher was able to receive feedback about her English proficiency in class as well as her teaching from the Japanese teacher who instructed her in the program. The advice provided by the junior high school teachers was as follows:

1. The student teachers' English proficiency
  - acquire English intonation well enough to teach
  - improve English pronunciation
  - improve English grammatical knowledge
  - practice everyday conversation in English
  - practice expressing oneself in English
  - practice before class in order to reduce the number of grammatical errors
2. The student teachers' teaching performance
  - use less Japanese
  - praise the junior high school students in English more frequently
  - relax and have more confidence when speaking English
  - do not be afraid of making mistakes because everyone makes mistakes
  - use language activities such as games more often
3. The junior high school teachers pointed out mistakes made by the student teachers, such as many grammatical errors in the exercises that they had prepared for classroom activities, ungrammatical sentences while instructing students in English such as "Are you finish?" and omission of "a" and "an" in speaking and writing.

Problems 1, 2, and 3 were observed on the videotape. The junior high school teachers' advice concerning the participants' English proficiency corresponded with the problems that the participants themselves identified while teaching. Although the advice is useful, it would have been more helpful if the problems had been stated in more specific terms; problems should have been pointed out in reference to the specific purpose of the activity and where the problem occurred in the teaching procedures. The junior high school teachers are competent enough to observe the student teachers' use of English in the classroom, but they are too busy to be trained to observe the student teachers' teaching performance and provide systematic feedback. A feedback checklist should be made as it would allow the junior high school teachers to provide higher quality feedback concerning student teachers' use of English and their teaching performance.

## APPENDIX D

### GRADE LEVELS AND PERIODS THE STUDENT TEACHERS TAUGHT

Student	Periods taught for 7 <sup>th</sup> graders (1 <sup>st</sup> year)	Periods taught for 8 <sup>th</sup> graders (2 <sup>nd</sup> year)	Periods taught for 9 <sup>th</sup> graders (3 <sup>rd</sup> year)
Student 1	12	0	0
Student 2	0	0	22
Student 3	0	4	10
Student 4	11	0	0
Student 5	0	12	0
Student 6	11	0	0
Student 7	0	3	8
Student 8	0	12	0
Student 9	0	6	4
Student 10	3	0	0
Student 11	0	13	0
Student 12	11	1	3
Student 13	0	0	11
Student 14	18	0	0
Student 15	20	0	0
Student 16	0	10	0
Student 17	0	14	0
Total	86	75	58
Average	5.0	4.4	3.4

## APPENDIX E

### SEVENTEEN EXAMPLES OF THE STUDENT TEACHERS' TEACHING

As introduction to each student teacher's performance, I first refer to the lesson number, the name of the textbook, the unit, the subunit, the style of text in the subunit, the type of the class (one period type or two period type), and the student number (the numbers are introduced in Appendix C). The style of text in the subunit means whether the text included a dialogue or not. Next, I introduce the content of each text. The title is placed in parentheses. "Introduction of text" (Text) introduces the outline of a whole text in each section, but, in one period, only one passage in the text is introduced.

Next, the student teachers' reading teaching is placed into one of four categories: their view of the text in the section, pre-reading activities, while-reading activities, and post-reading activities. In "view of the text," I do not introduce the outline the student teachers made or structural materials they wanted to teach through the text; instead, I report what the student teacher intended to lead their junior high school students to think about in relation to the following questions: What is the topic?; Why is the topic being written about?; How is this topic being written about?; What other ways of writing about the topic are there?; Who is the author writing to? (Wallace, 1990).

"Oral introduction of new words" mean that students practiced pronunciation of new words and the students answered with the words' Japanese meanings and/or the student teacher explained the Japanese meanings. "Oral introduction of the text" means that the student teacher reads aloud the text or memorized the text, and introduced the text orally, showing some pictures related to the content of the text. "Reading aloud practice" means that the students practice reading the text aloud, repeating a sentence/a long phrase after the student teacher reads it aloud. "Model reading" means that the students listened to the student teacher read or they listened to a tape. "Individual reading" means that the students were given a certain amount of time to practice reading aloud. "Role play" means that the students practiced a dialogue, taking the role of characters in the text. "T or F test" means that after listening to the introduction once or twice, the students were given about five questions to guide their listening, then they listened again to the introduction, and answered the questions in order for the student teacher to make sure that they understood the text. "T or F test" was sometimes used by the student teacher to make sure that the students had understood the main point in the text after reading. "Translation paper" means that the student teacher gave a complete Japanese translation of the text, or an incomplete form with some blanks for the students to fill in.

Teaching 7th graders

A. Sunshine English Course 1, Program 2- 2, a dialogue, (S 14)

Text: (A welcome party) Emily comes to Japan and stays with a Japanese family; they have a welcome party for her.

No view of the text

Pre-reading: oral introduction of new vocabulary

While-reading: model reading, reading aloud practice, and individual reading

No post-reading

B. Sunshine English Course 1, Program 4-2, a dialogue (S1)

Text: (A dialogue on the international telephone line) Emily, who is staying in Japan, calls her parents in the United States.

View of the text: Students think about cultural differences between the United States and Japan, as the text introduces "the time-gap," and "the American school system."

Pre-reading: Oral introduction of new words, oral introduction of the text, and a 5-item true-false test

While-reading: Model reading and reading aloud practice

The student teacher explained about "the time gap" as the topic was about an international telephone call and then the students practiced some typical expressions for communication on the phone.

Post-reading: A role-play

C. Sunshine English Course 1, Program 4-2, a dialogue (S12)

Text: (the same as B)

View of the text: Students can think about cultural differences between the States and Japan, as the text introduces "the time-gap," and "the American school system."

Pre-reading: Introduce new vocabulary

The student teacher explained in English the situation where the dialogue in the text occurs, and "the time gap."

While-reading: The students opened their textbooks, completed the true-false test, and practiced reading aloud

Post-reading: A role-play

D. Sunshine English Course 1, Program 5-1, a dialogue (S15)

Text: (Emily's experience in Japan) Now Emily goes to Kyoto with her Japanese family and learns about Japanese old customs and traditions there.

View of the text: Students can think about their own culture and cultural differences between the United States and Japan.

Pre-reading: Introduce new vocabulary, oral introduction of the text and a T or F test

While-reading: Translation of the text into Japanese, model reading, and reading aloud practice

No post-reading activities

E. Everyday English 1, Lesson 4-2, a dialogue (S4)

Text: (Introducing people) Koji, Bekky, Megumi, and Dick meet each other for the first time and introduce themselves to each other.

No view of the text

Pre-reading: Introduce new vocabulary and practice the usage of *he* and *she*.

While-reading: There are pictures of the characters in the text; the students determine which picture is being introduced while reading; reading aloud practice.

Post-reading: Using expressions in the text, the students introduced themselves in English.

F: New Horizon English Course 1, Lesson 3-2, a dialogue (S10)

Text: (Introducing people) Kumi and Ken visit Mike; there they talk about things in Mike's room, their friends, and their families.

View of the text: Students think about communication.

Pre-reading: Introduce new vocabulary, listen to the tape, and administer a true-false test.

While-reading: reading aloud practice and translation of the text.

Post-reading: a role-play.

G: New Crown English series 1, Lesson 6, a dialogue (S6)

At this junior high school, in the first term of the 7th grade, the students do not start to learn to read and write yet.

Teaching reading to 8th graders

H: Sunshine English course 2, Program 4-2, no dialogue, One period type, (S16)

Text: (Soccer in Brazil) A journalist goes to Brazil and describes how soccer is deeply rooted in Brazilians' life.

No view of the text

Pre-reading: Introduce new vocabulary, oral introduction of the text, and a true-false test

While-reading: Reading aloud practice and the student teacher provided background knowledge about the content of the text.

Post-reading: The students did a role-play, and then memorized the text.  
Translation sheet

I: New Horizon English Course 2, Lesson 4-1, a dialogue, Two period type (S9)

Text: (Paula's Summer Vacation) Paula and her family go to London during her summer vacation, and she talks about sight-seeing spots.

View of the text: The text is about sight-seeing spots in London, but through the text, I hope to lead the students to become interested in the United Kingdom and know more about the country.

Pre-reading: introduction of new words, sometimes with explanation about the words when the student teacher thought it necessary or when the usage of a new verb was important.

While-reading: model reading and reading aloud practice. The students were given a few minutes to read silently and then the student teacher asked several questions in English and/or Japanese. After finding which part of the text was difficult for the students to understand, the student teacher translated part of the text and explained grammar, contexts, and so on.

Post-reading: a role-play: Some pairs were asked to present their role-play in front of the other classmates.

J. Sunshine English Course 2, Program 3-2, no dialogue, Two period type (S5)

Text: (Interesting things and places in Australia) Kumi has been studying in Washington D.C., and one of her teachers, Mr. Wood talks about his trip in Australia one week after he came.

View of the text: Students get interested in Australia.

Pre-reading: introduction of new vocabulary

There are two pictures and one illustration related to the context of the text, which is a dialogue between Kumi and Mr. Wood. One of the pictures is a koala, another shows a girl holding a koala, and the illustration is a koala. The teacher asked several questions in English or Japanese about koalas, while pointing at these pictures.

While-reading: The students listened to the tape once, and then read the text silently. The teacher asked several questions about the content of the text and checked if the students understood the text.

No post-reading activities

K. Sunshine English Course 2, Program 3-4, a dialogue, One period type (S8)

Text: The same as J

View of the text: This is the first section in the textbook that introduces a foreign country, so I hope the text leads the students to become interested in other foreign countries as well as Australia and to try to understand non-Japanese persons.

Pre-reading: oral introduction of the text and introduction of the new vocabulary

While-reading: reading aloud practice. The students were instructed to find the new sentence pattern in the text. The teacher explained some new phrases.

No post reading: only translation sheet

L. Sunshine English Course 2, Program 3-4, a dialogue, One period type (S11)

Text: The same as J

View of the text: Japanese students know much about places in Australia but not much about its history. The text may lead them to take an interest in it and also have attitude to understand people abroad.

Pre-reading: introduction of the new words, introduction of the text, and a true-false test

While-reading: to practice reading aloud and to find and underline the new sentence pattern in the text

No post-reading: Only translation sheet

M. New Crown English Series 2, Lesson 4-1, no dialogue, One period type (S17)

Text: (Gestures and communication) The text introduces gestures as a communication tool and also describes cultural differences in gestures.

No view of the text

Pre-reading: In the oral introduction of the text, the student-teacher read the text aloud while showing pictures related the content of the text. When introducing new words, she focused on their pronunciation and their Japanese meanings.

While-reading: The student-teacher asked questions related to the text, such as, "What do people use for communication?". The students practiced reading aloud after listening to the model reading of a native speaker of English on the tape.

No post-reading

Teaching reading to 9th graders

N. New Crown English series 3. Lesson 4-3, a dialogue. Two period type (S3)

Text: (School trip to Hiroshima) Ken goes to Hiroshima on his school trip and writes a letter from Hiroshima to a friend. Ken writes about how he feels about peace.

View of the text: The students can think about the danger of atomic bombs through reading the text, and also take interest in Hiroshima, as they are going on a school trip to Hiroshima as well as Kyoto.

Pre-reading: The students first read in Japanese about the atomic bomb dropped on Hiroshima and then discuss the topic. (The students were interested in the topic, partly because they were going to Hiroshima on their school excursion.). In the oral introduction of the text, the student-teacher introduced the content of the text in simple English and asked questions in English sometimes.

While-reading: The students practiced reading aloud. First, the teacher read slowly, and gradually read at a higher speed.

Post-reading: a role-play and translation sheet. The students tried to memorize the text.

O. New Crown English series 3. Lesson 4-3, a dialogue. One period type. (S7)

Text: The same as N

View of the text: The students will understand how valuable peace is through reading the text.

Pre-reading: When introducing new words, the student-teacher focused on the pronunciation and Japanese meanings of the words. Oral introduction of the text.

While-reading: The students listened to the teacher's model reading and repeated after her. The teacher asked the students to translate the text into Japanese.

No post-reading

P. New Total English Book 3, Lesson 3-1, a dialogue, One period type (S2)

Text: the text is part of a Japanese fantasy translated into English; Giovanni had had a friend called Campanella, who died after saving Giovanni from drowning. One day, Giovanni got on a ghost train for the other world and met Campanella there.

View of the text: The text may lead the students to become interested in Japanese literature.

Pre-reading: introduction of new words. The students were told to find the new structure pattern in the text.

While-reading: The students listened to the teacher's model reading twice, and the tape twice, and then practiced reading aloud. The teacher asked the students to translate the text and then the teacher provided grammatical explanations.

Post-reading: The teacher read five sentences orally and the students responded by saying "true" if the meaning of each of these sentences corresponded with the text and "no" if not.

Q. New Total English Book 3, Lesson 3-2, no dialogue, One period type (S13)

Text: The same as P

View of the text: The text may lead the students to think about friendship, courage, and love.

Pre-reading: introduction of the new words, explanation of the usage of the new auxiliary *had better*, and oral introduction of the text

While-reading: The teacher asked the students to say the Japanese meaning of each sentence in the text. The students listened to the teacher reading aloud the text once, and then read the text twice following the teacher.

No post-reading

APPENDIX F

THE RASHOMON EFFECT (JAPANESE VERSION)

実習生 授業 内容	A	A	B	B	C	C	D	D	E	E
前の授業の復習										
インターアクションゲーム の評価と確認										
新しい表現の紹介										
新出単語の紹介・練習										
新しい文型の文法説明と練習										
本文内容理解のための質問										
新しい文型を用いた言語活動の紹介										
教科書本文の暗唱練習										
インターアクションゲーム の取り組み方説明のデモン ストレーション										
教科書本文の日本語訳										

## APPENDIX G

### THE TEACHING SKILL QUESTIONNAIRE (ENGLISH VERSION)

#### SELF-ASSESSMENT

1. Greet in English

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

2. Answer when a student asks about the Japanese meaning of an English word in the textbook

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

3. Translate reading in the textbook into Japanese

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

4. Explain a new structure in Japanese

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

5. Provide background knowledge of the reading in the textbook

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

6. Notice serious mistakes in the students' English and correct them immediately

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

7. Do a good model reading

1 = easy                      2 = fairly easy                      3 = a little easy                      4 = not difficult nor easy  
5 = a little difficult                      6 = fairly difficult                      7 = difficult

8. Answer questions about grammar

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

9. Instruct students in Classroom English

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

10. Make an oral introduction of a new sentence pattern in English

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

11. Provide good examples of the new structure

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

12. Ask questions in English about the text and have the students try to answer in English

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

13. Make an authentic and interesting language activity such as an information gap task

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

14. Introduce the context in your own words in English

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

15. Prepare for the next team-teaching class by discussing it with an ELT in English

1 = easy	2 = fairly easy	3 = a little easy	4 = not difficult nor easy
5 = a little difficult	6 = fairly difficult	7 = difficult	

**APPENDIX H**  
**THE REVISED COLT**  
**BLOCK A**

Block B - Block C. These are always some boxes checked with a circle or description of activities, but please change to no circle or double circle if you disagree. A double circle is used when certain activities are very eminent.

Time	Procedure / Topics of the class	Focus of the lesson	Teacher's activities	Students' activities	ST	ST
15:30 - 15:47 Activity 3	Teaching materials		Student activities: Student modality: Writing, reading, speaking, listening Participant organization: class, individual Sustained speech: minimal, ultra-minimal Interlocutor: peers, teacher Use of language: L1, L2 Teacher's questions: genuine, pseudo Purpose of use of the language: content, socializing or instruct Use of language: L1, L2 Interlocutor: class or group, individual	Teacher's activities: Teacher's questions: genuine, pseudo Purpose of use of the language: content, socializing or instruct Use of language: L1, L2 Interlocutor: class or group, individual		
15:47 - 16:00 Activity 4	Grammatical Explanation ; Introduction of interactive activities	culture and/ or etc discourse and/ or sociolinguistics function grammar, vocabulary, and/ or pronunciation at the word level and/ or sentence level				
					paced by ST related with Ss' everyday life	Text-book New Sentence Pattern
					00:08:57 Activity 1	00:09:00 - 00:14:00 Activity 2



Block C

Block C : There are already some boxes checked with a circle or description of activities, but please change to no circle or a double circle if you disagree. A double circle is used when certain activities are very eminent.

		Student modality					
Students' Interactive activities	Sustained speech	minimal				/	
		ultra-minimal				/	
	interaction	fixed				/	
		To some extent free				/	
		free				/	
	vocabulary	words to choose/ free				/	
		fixed				/	
	Sentence pattern	not fixed				/	
		fixed				/	
	Students' activities	Participant organization	class/ individual	class	individual	class	
		Interlocutor	peers teacher				/
		Use of language	L1				
L2							
Teacher's activities	Sustained speech	minimal					
		ultra-minimal					
	Teacher's questions	genuine					
		pseudo					
	Purpose of use of	content	content				
	Use of language	L1					
L2							
Interlocutor	class or group		○	/	○		
	individual		○	/	○		
Focus of the lesson	culture and/ or etc discourse and/ or sociolinguistics		○	○	○		
		function	○	○	○		
	grammar, vocabulary, and/ or pronunciation at the word level and/ or sentence level	○	○	○	○		
Teach-in mate rails	Text, Stg prepared by ST						
Topics	Related with Ss' everyday life						
Procedures of the class			Explanation of the Game demo	Students' Interactive activities	evaluation		
Time			26:44-31:55 Activity 5	31:55-34:20 Activity 6	34:23-36:00 Activity 7		





**APPENDIX I**  
**THE STUDENT TEACHERS' VIDEOTAPED INSTRUCTION**  
**(ENGLISH VERSION)**

**BLOCK A**

1. Is it useful if a teacher reviews the class in English trying not to use Japanese?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

2. The student teacher on the screen tries to review the class in English trying not to use Japanese. Is she effective when her English and her teaching technique is also taken into consideration?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

3. Is it useful if a teacher introduces a new sentence pattern in English trying to lead students to discover a new rule?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

4. The student teacher tries to introduce a new sentence pattern in English in order to lead the students to discover a new rule. Is she effective when her English and her teaching technique are taken into consideration?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

In these two scenes, the students just listened except some occasions. Is this useful for review? Is this useful for introducing a new sentence pattern?

5. Review

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

6. Introduction

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

## BLOCK B

7. Is it useful to explain about English grammar in Japanese?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

8. The student on the screen explains about new items of English grammar. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

9. In order to learn a new sentence pattern, is this useful for students to repeat the new pattern after the teacher?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

10. Does the student teacher on the screen let her students repeat after her. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

11. Generally speaking, in order to learn a new sentence pattern, is it useful for students to practice such language activities as information gap/interview tasks?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

12. Is the interview game the student teacher on the screen introduces effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

13. Is it useful for students to translate a Japanese sentence into English using the new sentence pattern?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

14. The student teacher let the students translate several isolated Japanese sentences into English. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

## BLOCK C

15. Before the language game, is it useful to remind students about eye-contact, intonation, and aspects of what is involved in face-to-face communication as well as to explain how to play the game? After the language game, is it useful for students to check their eye contact, intonation, and other aspects of face-to-face communication?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

16. Is the student teacher's instruction before the game and after the game effective? (She reminds the students of eye-contact, intonation, etc. and explains how to play the game. After the language game, she asks the students to check their eye contact behavior, intonation, etc.)

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

17. Is it more effective if language games such as the interview game are authentic?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

18. Is the language game on the screen authentic?

1 = Not authentic    2 = Not very authentic    3 = between '2' and '4'    4 = Authentic    5 = Very authentic

19. Is it useful if a language game is interesting to students?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

20. Can the game on the screen be interesting to students?

1 = I strongly disagree    2 = I disagree    3 = between '2' and '4'    4 = I agree    5 = I strongly agree

21. Which type of language activities are most appropriate for junior high school students?

1. Performing memorized dialogues
2. Contextualized drills—everything is controlled by the teacher
3. A language game in which students are supposed to use a new sentence pattern. They can choose any words they like.
4. Cued dialogues
5. Improvisation

22. Which type of language activity do you think the game on the screen is?
1. Memorized dialogue
  2. Contextualized drill—everything is controlled by the teacher
  3. A language game in which students are supposed to use a new sentence pattern. They can choose any words they like.
  4. Cued dialogue
  5. Improvisation

23. Overall, is the game on the screen planned by the student teacher effective?

1 = Not effective      2 = Not very effective      3 = between '2' and '4'      4 = Effective      5 = Very effective

### BLOCK D

24. Is it useful to introduce new words with flash cards, check the students' pronunciation, and the meaning of the words before beginning to read a text?

1 = Not useful      2 = Not very useful      3 = between '2' and '4'      4 = Useful      5 = Very useful

25. Is the student teacher effective?

1 = Not effective      2 = Not very effective      3 = between '2' and '4'      4 = Effective      5 = Very effective

26. Is it useful to introduce the contents of the text orally with picture cards before starting to read the text?

1 = Not useful      2 = Not very useful      3 = between '2' and '4'      4 = Useful      5 = Very useful

27. The student teacher introduces the text orally with picture cards before starting to read the text. Is she effective?

1 = Not effective      2 = Not very effective      3 = between '2' and '4'      4 = Effective      5 = Very effective

28. Is it useful to make students aware of the new sentence pattern and new aspect of English grammar while reading?

1 = Not useful      2 = Not very useful      3 = between '2' and '4'      4 = Useful      5 = Very useful

29. The student teacher let the students underline sentences that include the new sentence pattern so that they will be aware of the pattern in the text. Is she effective?

1 = Not effective      2 = Not very effective      3 = between '2' and '4'      4 = Effective      5 = Very effective

30. Is it useful to ask questions in Japanese about the text?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

31. Is it useful to ask questions in English about the text?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

32. Overall, is the student teacher effective in asking questions either in English or Japanese?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

33. Is it useful to have students practice reading the text aloud?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

34. Is the student teacher's reading aloud instruction effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

35. This is not what the student teacher on the screen is doing. Is it useful to give students background knowledge of the text and/or activate their background knowledge before starting to read the text?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

36. This is not what the student teacher on the screen is doing. Is it useful to let students grasp the points of the text before reading intensively?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

## **BLOCK E**

37. Is it useful to have students memorize the text as part of the post-reading activities?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

38. The student teacher on the screen encourages the students to recite what they have memorized by giving hints, paraphrasing, correcting, and completing the sentences. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

39. Is it useful to have students translate the text into Japanese?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

40. The student teacher had the students translate the text into Japanese. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

41. Is it useful to talk about why the text is worth reading?

1 = Not useful    2 = Not very useful    3 = between '2' and '4'    4 = Useful    5 = Very useful

42. The student teacher explains in Japanese why the text is worth reading as part of post-reading activities. Is she effective?

1 = Not effective    2 = Not very effective    3 = between '2' and '4'    4 = Effective    5 = Very effective

**APPENDIX J**  
**LETTER OF ACCEPTANCE**

1. This study is being conducted so that college teachers and student teachers can receive more specific feedback from junior high school teachers who supervise student teachers in their teaching practicum.

2. Another purpose of this study is to compare the viewpoints of three groups concerning teaching English (student teachers, college teachers, and junior high school teachers).

I agree to take part in this study and to have my responses used in the study.

Name:

Signature:

The person collecting and using the data is Sadayuki Mitsuo.

The Department of British and American Culture

The Faculty of the Modern Culture

Tokyo Junshin Women's College