Assessing Student Evaluation of English Activities Using the COLT Observation Scheme

Yoshiki Yokoyama (Hokkaido University of Education)
Akinobu Shimura (Asahikawa Jitsugyo Senior High School)
Michelle La Fay (Hokkaido University of Education)
Hiroki Ishizuka (Hokkaido University of Education)
Yasushi Kawai (Hokkaido University)
Chikako Aoki (Hokkai-Gakuen University)

Abstract

In this study, we used a mixed method (Dörnyei, 2003) so that the student evaluation was complemented by classroom observation data from the Communicative Orientation of Language Teaching (COLT) Observation Scheme (Spada & Fröhlich, 1995). To collect the data for this study, we conducted a questionnaire survey of classroom activities, and observed six classes at one university to investigate whether the students' sense of preference or effectiveness reflect the time spent on an activity and/or the frequency of an activity and whether the activities the students rated more preferable or effective have some features in common in terms of the COLT Observation Scheme. The results show that some evaluation scores reflect the observation data, and such findings were seen more in group work than in individual activities. With respect to COLT features, topic control and student modality have an impact on student evaluation.

1. Introduction

1.1 Literature Review

Recently Communicative Language Teaching (CLT) has come to be seen as one of the best methodologies for English learners. In communicative English classes, most teachers provide many kinds of activities for their students. There are many arguments about the advantages and disadvantages of the types of activities CLT has used. To investigate student attitudes toward class activities, many studies employed a questionnaire method. Its advantage is the efficiency in administration. One can gather a huge amount of information in a relatively short period of time. However, this kind of student evaluation has some limitations. One is that students do not always provide true answers about the activities they have encountered. To compensate for the drawbacks of the questionnaire data, Dörnyei (2003) suggests a mixed method in which a questionnaire is used in conjunction with qualitative measures. In this study, student evaluation will be complemented by one of these qualitative measures, the Communicative Orientation of Language Teaching (hereafter COLT) Observation Scheme presented by Spada and Fröhlich (1995).

The COLT Observation Scheme is the most popular classroom observation scheme to measure communicative orientation in the classroom and to investigate the relationships between teaching and learning. The COLT Observation Scheme consists of two parts: COLT Part A which focuses on describing features of activities (henceforth COLT features) that include participant organization, content, student modality and materials, and COLT Part B which analyzes the communicative features of verbal interactions between teachers and students or students and students during classroom activities (Fröhlich, Spada & Allen, 1985; Spada & Fröhlich, 1995).

Past studies concerned with the COLT Observation Scheme have

been done to identify the communicative orientation of language classes. Some of the studies used the categories of the COLT Observation Scheme as dependent variables to measure the features that occurred in language classes, and others used them to identify the independent variables that exist in language classes. The former type includes Sugita (2002), Yoneyama, Oominato, Suda, Maeno, and Yoshida (2003), Anno (2003), and Sugino, Koga, and Kawashima (2004). Sugita and Yoneyama, et al. used the COLT Observation Scheme to observe junior high school classes conducted by the same teacher, and different teachers respectively. The results of both studies clearly revealed different features of the classes, and thus confirmed the validity of the COLT Observation Scheme as an observation tool. Anno observed elementary school English classes, and Sugino did longitudinal research on a kindergarten English class. They both obtained clear COLT features distinguishing the classes in different periods and grades. Sugino concluded that the COLT Observation Scheme is useful for clarifying the change of activity contents in a longitudinal study. Anno suggested that the COLT Observation Scheme reflects the aim of activities and assists teachers in making syllabuses and teaching plans appropriate for each grade, thus devising and improving the way of teaching and engaging pupils.

The studies which used the COLT Observation Scheme to identify some features of independent variables include Yamada (1996), Panova and Lyster (2002). Yamada used the COLT Observation Scheme to identify the features of three different elementary school lessons in order to examine whether the subjects' comprehension differs with the amount of L2 used in class, and suggested that L2 use would be recommended in elementary school classes. Panova and Lyster examined the relationship between different types of corrective feedback by a teacher and types of uptake by 25 learners in an adult ESL class in Canada. They used the COLT Observation Scheme to clarify instructional features of the class

and found that 90% of the class time was spent for oral exchanges between a teacher and the students, which indicated a high degree of communicative orientation.

Many of the past studies claimed that the COLT Observation Scheme is a useful observation tool to identify independent variables and measure features of a language class. However, none of the studies have attempted to conduct a mixed method in which the COLT Observation Scheme is used to investigate questionnaire data. The present study thus attempts to reveal whether COLT coding can reflect learners' perception of the activities gathered in the questionnaire data.

1.2 Research Questions

A questionnaire survey is often administered to investigate language classes, attempting to reveal what activities are preferred or perceived effective. However, the data collected from the questionnaire should be carefully investigated, since they may not reflect what actually occurred in the class. To reduce such shortcomings, this study proposes the use of the COLT Observation Scheme which observes classroom activities in terms of the time spent on an activity, the frequency of an activity, and the COLT features an activity is coded for. The study thus aims primarily to answer the following research questions.

1. Does the students' sense of preference or effectiveness reflect the time spent on an activity and/or the frequency of an activity?
2. Do activities the students rated more preferable or effective have some features in common in terms of the COLT Observation Scheme?

2. Method

To answer the questions, we conducted a questionnaire survey of classroom activities and classroom observation of six classes at one university.

2.1 Participants

This experiment was conducted at one university in Japan in six English classes: three freshman speaking classes and three sophomore intensive speaking classes. Three different teachers were observed and each taught one speaking class and one intensive speaking class. The number of survey participants is shown in Table 1.

Table 1.

Number of Participants

Teachers	Class type	# of participants
A	Speaking	15
	Intensive Speaking	8
B	Speaking	20
	Intensive Speaking	8
C	Speaking	20
	Intensive Speaking	8

Students were asked to sign a consent form stating that they were willing to participate in our study. Students were informed that all results would be confidential and neither their names nor other private information would be released. The teachers also signed consent forms stating they agreed to class observation.

2.2 Materials and Procedures

There are two parts to the experiment: a questionnaire survey and class observation according to the COLT Observation Scheme. Preparation for this study began with the development of a preliminary questionnaire gauged for university students studying English as part of the required curriculum. The preliminary questionnaire was written in Japanese with an English translation and was given to the participants

one week before class observation. It was designed to gain information about the activities in class.

The preliminary questionnaire consisted of 13 questions about three kinds of participant organization: group, teacher-centered, and individual[1]. First, students were asked about activities in each kind of participant organization, and then they were asked to write in detail about the activities done in class. Students were also asked if the teacher used Japanese in class and how the teacher corrected errors. In addition, information about topics of conversation and the use of textbooks was obtained. The results of the preliminary questionnaire were tallied and analyzed in order to formulate the survey.

The actual questionnaire survey given after class observation was designed to assess the students' preference and perceived effectiveness of class activities and tasks. The survey centered on specific activities that the students had identified as preferable or effective and consisted of 26 questions. For each of the three kinds of participant organization: group, teacher-centered, and individual, students were first asked if they liked the activity, and then if they thought the activity was effective, and finally, exactly what kind of activity they had participated in.

Students were given several choices for each category and multiple answers were acceptable (See Table 2). Results of the survey were analyzed obtaining percentages of preference and effectiveness for each question and these results were then collated with the results of COLT Part A coding in order to analyze the classes.

The second part of this study consisted of the actual class observation which was carried out using both IC recorders and video cameras in order to capture both conversation and the flow of the classes. Video cameras were set in the front and back of class. IC recorders were distributed among the students to record conversation between the students. Two researchers were delegated to each class to ensure smooth operation of

equipment and to conduct the survey.

Table 2.

Activities by Participant Organization

Group activities	Teacher-centered activities	Individual activities
Reading dialogue	Instructor assessing	Solving problems
Role-playing	pronunciation of the	Reading aloud
Talking about a	student(s)	Silent reading
particular topic	Repeating after the	Consulting a
Conversation	teacher	dictionary
Solving problems	Providing answers	Writing
and checking the	Presenting language	
answers	forms	
	Questions and answers	
	Listening practice	

After completion of class observation, class conversations were transcribed by watching the videos and listening to the recordings of the classes. The conversations were transcribed using AS-units (Foster, Tonkyn, & Wigglesoworth, 2000) to identify separate units of speech. After the transcripts were completed, COLT Part A coding was carried out. Each researcher coded one specific class. Finally, the questionnaire survey and the results of COLT coding were analyzed together to assess whether the students' preference and effectiveness were related to the results of the COLT coding.

In order to facilitate the comparison of results among the classes, each researcher compiled tables of data in order to show the relationships between students' preference/effectiveness and time/frequency or COLT features coded for each activity. The following table is an excerpt from Intensive B speaking class analysis and shows how each researcher

organized their results. In Table 3, a positive correlation between frequency and effectiveness can be seen in "presenting language forms" where the activity occurred four times and students rated this activity as effective (66%). However, it can conversely be seen that the "listening practice" with a time of nine minutes six seconds and a frequency of three was rated fairly preferable (33%), but was not assessed to be effective by the students.

Table 3.

Evaluation Scores and Coding Results of Teacher-centered Activities in Intensive B

Activities	Evaluation		Time	Freq.	Participant Organization	Content
	Pref.	Effect.				
Presenting language forms	44.4%	66.7%	2:53	4x		
Explaining about discourse 1			1:11	2x	T⇔ S/C[a]	Discourse
Explaining about discourse 2			1:01	1x	T⇔ S/C	Discourse
Explaining about vocabulary			0:41	1x	T⇔ S/C	Form/V[b]
Listening practice	33.3%	11.1%	9:06	3x		
Listening to CD 1			5:59	1x	T⇔S/C	Narrow
Listening to CD 2			1:08	1x	T⇔S/C	Narrow
Explaining about a personal experience			1:59	1x	T⇔S/C	Narrow

Note. [a]T⇔S/C means that the teacher interacts with the whole class or individual students.

[b]Form/V means primary focus of instruction is on form, especially vocabulary.

With respect to COLT features, content varies with the activities;

"presenting language forms," which focuses on language (vocabulary or discourse) was rated more effective than "listening practice," which deals with a topic referring to the students' immediate environment and experience. After each researcher created this kind of table for their data, they then shared the tables and results with the other researchers. Tables 4 through 10 were then prepared for the current analysis to show the data under each activity type, and how the evaluation scores reflected time/frequency or COLT features.

3. Results

3.1 Time and Frequencies

Language teachers usually organize their classes so as to make important activities longer and more frequent. However, the students do not always prefer or perceive effective an activity that is longer and/or more frequent. To find the gap between a teacher's intension and the students' perceptions (Research Question 1), attitudinal data toward class activities were investigated in terms of the time spent on an activity and the frequency of an activity. Table 4 shows the time spent on the activities and the preference and effectiveness scores gained from student evaluations of activities. In Speaking A, there were two activities in group work, "talking about a particular topic" and "conversation." The time spent on the activity was eighteen minutes fifty seconds and three minutes twenty-five seconds respectively. In this class, effectiveness scores increased with the time spent on the activity in group work. "Talking about a particular topic" was rated effective by 53.3% of the students and "conversation" by 35.0%. Preference scores in group activities, however, decreased with the time spent on them. "Talking about a particular topic" was rated as preferred by 13.3% of the students and "conversation" by 40.0%.

In Speaking B, preference ratings increased with the time spent on the activities in group work. "Conversation" was rated effective by 45.0% of the students and "talking about a particular topic" by 30.0%. Effectiveness ratings in group work, however, decreased with the time spent. "Conversation," which lasted ten minutes twenty-eight seconds was rated 35.0%, while "talking about a particular topic," which lasted three minutes three seconds was rated 55.0%.

In individual work, both preference and effectiveness scores decreased with the time spent on the activities. "Writing practice," which lasted eight minutes thirty-five seconds, was rated preferred by 10.0% of

Table 4.

Time and Preference/effectiveness Within Activity Types

Class	Participant Organization	Activity	Time	Pref.	Effect.
Speaking A	Group work	Talking about a particular topic	18'50"	13.3 %	53.3 %
		Conversation	3'25"	40.0 %	35.0 %
Speaking B	Group work	Conversation	10'28"	45.0 %	35.0 %
		Talking about a particular topic	03'03"	30.0 %	55.0%
	Individual	Writing practice	08'35"	10.0 %	10.0 %
		Solving problems	01'38"	70.0 %	85.0 %
Intensive A	Teacher-centered	Providing answers	17'43"	0.0%	25.0 %
		Presenting language forms	09'31"	87.5%	87.5 %
	Individual work	Solving problems	09'12"	75.0%	100.0 %
		Listening practice	02'51"	37.5%	37.5 %

Table 4. *(continued)*

Class	Participant Organization	Activity	Time	Pref.	Effect.
Intensive B	Group work	Conversation	12'10"	88.9%	77.8 %
		Talking about a particular topic	09'46"	44.4%	44.4 %
		Reading dialogue	02'19"	22.2%	0.0 %
Intensive C	Group work	Talking about a particular topic	48'20"	62.5 %	87.5 %
		Solving problems and checking the answers	01'13"	37.5 %	0.0 %
	Teacher-centered	Questions and answers	06'03"	62.5 %	62.5 %
		Listening practice	01'07"	25.0 %	25.0 %

the students and effective by 10.0 %. On the other hand, "solving problems," which lasted only one minute thirty-eight seconds, gained much higher scores both in preference and effectiveness.

In Intensive A, both preference and effectiveness scores in individual work increased when more time was spent on the activities. However, such positive relations were not seen in teacher-centered activities; "providing answers" was rated much lower than "presenting language forms," although the former lasted longer than the latter.

In Intensive B, a positive relation was seen in group work, and in Intensive C, such relations were observed both in group work and teacher-centered activities. The findings that preference and effectiveness scores increased with the time spent on the activities were only seen in

Intensive classes, not Speaking classes. Overall, more findings were seen in group work and teacher-centered activities.

Table 5.

Frequencies and Preference/Effectiveness Within Activity Types

Class	Participant Organization	Activity	Freq.	Pref.	Effect.
Speaking A	Group work	Talking about a particular topic	4x	13.3 %	53.3 %
		Conversation	1x	40.0 %	35.0 %
Intensive A	Teacher-centered	Presenting language forms	8x	87.5 %	87.5 %
		Providing answers	4x	0.0%	25.0 %
Intensive B	Teacher-centered	Presenting language forms	4x	44.4 %	66.7 %
		Listening practice	3x	33.3%	11.1%
Intensive C	Group work	Talking about a particular topic	8x	62.5 %	62.5 %
		Solving problems	1x	37.5 %	0.0 %
	Teacher-centered	Questions and answers	2x	62.5 %	62.5 %
		listening practice	1x	25.5 %	25.0 %

Table 5 shows how frequencies of activities reflect on the preference and effectiveness scores gained from student evaluation of the activities. In Speaking A, "talking about a particular topic," which was repeated four times, was rated effective by 53.3% of the students, while "conversation," which was implemented only once was rated effective by 35.0%. In this class, effectiveness scores increased when more frequent group work was

conducted. Preference in group work, however, decreased with the frequency. "Talking about a particular topic" was rated as preferred by 13.3% of the students and "conversation" by 40.0%. As is seen in the previous section, more positive relations were seen in intensive classes. In Intensive A, B and C, both preference and effectiveness scores increased when more frequent activities were conducted. Such positive relations

Table 6.

Time and Preference/Effectiveness Beyond Activity Types

Class	Participant Organization	Activity	Time	Pref.	Effect.
Intensive B	Group Work	Conversation	12'10"	88.9 %	77.8 %
		Talking about a particular topic	09'46"	44.4 %	44.4 %
	Teacher-centered	Listening practice	09'06"	33.3 %	11.1 %
		Presenting language forms	02'53"	44.4 %	66.7 %

were seen in group work and teacher-centered activities, but not in individual activities.

In the previous section, we analyzed the activities grouped into either teacher-centered, group work or individual activities. That is, we have limited the domain of analysis to within the activity types. We will then extend it to beyond activity types. This usually presupposes more findings, but the results indicated fewer findings. Table 6 shows how the time spent on the activities reflects on the preference and effectiveness scores in beyond activity types. Only Intensive B had positive relations showing that the time spent may be one of the elements that influence preference.

"Conversation" which was the longest activity had the highest score both in preference and effectiveness, while the shortest activity, "presenting language forms," was rated second most effective. There was no strong tendency in other classes.

Table 7.

Frequency and Preference/Effectiveness Beyond Activity Types

Class	Participant Organization	Activity	Freq.	Pref.	Effect.
Intensive B	Teacher-centered	Presenting language forms	4x	44.0 %	66.7 %
		Listening practice	3x	33.3 %	11.1 %
		Providing answers	2x	0.0 %	11.1 %

Table 7 shows how frequencies of the activities reflect on the preference and effectiveness scores in beyond activity types. Moderate tendency was seen only in Intensive B, in which "presenting language forms" that occurred four times was rated most effective over the other activities that occurred two or three times. There was no other strong tendency in other classes.

3.2 COLT Features

Table 8 shows how evaluation scores obtained from the questionnaire survey reflect coding results of three kinds of activities. To address Research Question 2, activities are grouped into either group work, teacher-centered or individual activities to investigate how COLT features can account for questionnaire scores. Table 8 indicates that more findings were seen in group work and teacher-centered activities than in individual activities. With respect to other COLT features, topic control and student modality have an impact on student evaluation scores. In

Speaking A and Intensive B classes, group activities with more content control on the part of the students tended to be preferred, while moderate content control was preferred over free conversation in Intensive A, showing excessive student control was not always favored.

Multimodal activities were positively evaluated in two of the sophomore classes, Intensive A and Intensive C, while in Intensive B, reading activities were less preferred and perceived less effective. All these probably reflect the tendency that sophomore students preferred

Table 8.

Student Evaluation and COLT Coding Results Within Activity Types

Class	Participant Organization	Preference	Effectiveness
Speaking A	Group work	Activities with more content control on the part of the students tended to be preferred.	—
	Teacher-centered	Choral activities were preferred.	Choral activities were perceived effective.
Intensive A	Group	Moderate content control was most preferred, followed by free conversation, and activities with almost no options.	—
	Teacher-centered	Activities which required both listening and speaking were highly preferred.	Activities which required both listening and speaking were perceived as highly effective.

Table 8. *(continued)*

Class	Participant Organization	Preference	Effectiveness
Speaking B	Teacher-centered	Activities which required listening were highly preferred.	Effectiveness was perceived to be higher when the students were listening directly to the teacher as opposed to the CD.
Intensive B	Group work	Activities with more content control on the part of students were preferred. An activity requiring reading a longer text was less preferred.	Activities with more content control on the part of students were perceived effective. An activity requiring reading a longer text was perceived less effective.
	Teacher-centered	—	An activity which focused on discourse was perceived most effective.
Intensive C	Group work	Activities which required writing, listening and speaking were preferred.	Activities which required writing, listening and speaking were perceived highly effective.

Note. — indicates no relationship found. N/A means not applicable since there is not enough data for the analysis.

more communicative-oriented activities, which were also perceived effective. Even in Intensive B, an activity in which most of the episodes focused on discourse was perceived effective. On the other hand, choral activities were preferred and perceived effective in one of freshman classes, Speaking A. Single modality was positively evaluated in another freshman class, Speaking B. This may reflect the idea that simpler or mechanical activities were preferred and perceived effective by the freshman.

Table 9.

Student Evaluation and COLT Coding Results Beyond Activity Types

Class	Preference	Effectiveness
Speaking A	—	Activities with more content control on the part of the students were perceived effective.
Intensive A	—	Multiple modality increased effective scores
Intensive B	Activities that focus on discourse tended to be preferred.	Activities that focus on discourse tended to be perceived effective.
Speaking C	Content control from the instructor both in group work and in teacher-centered activities was preferred.	Content control from the instructor both in group work and in teacher-centered activities was perceived effective.

Table 8 limited the domain of analysis to within the activity types. Table 9, on the other hand, extended the scope, attempting to compare activities within each class. This usually presupposes more findings. Table 9, however, shows fewer findings than Table 8, indicating that

student evaluation will possibly offer more reliable results when we limit the domain of analysis to within activity types. With respect to COLT features, as seen in Table 8, topic control and student modality have an impact on student evaluation scores. In Speaking A, more content control on the part of the students was perceived effective, while content control from the instructor was preferred and perceived effective in Speaking C. Another COLT feature is multiple modalities which tend to increase effective scores. Such communicative orientation was also found in Intensive B, in which the students preferred and perceived effective the activities that focused on discourse.

4. Conclusion

This study employed a mixed method which used a questionnaire for student evaluation and classroom observation by the COLT Observation Scheme. We aimed to investigate positive relations between students' perception of preference/effectiveness regarding time/frequency and COLT features. This study also suggested the validity of student evaluation of class activities. There are two research questions: 1. Does the students' sense of preference or effectiveness reflect the time spent on an activity and/or the frequency of an activity? 2. Do activities the students rated more preferable or effective have some features in common in terms of the COLT Observation Scheme?

Regarding the first research question, according to the results of analysis within activity types, we were able to conclude to some degree that as the time/frequency of activities increased, the preference/perceived effectiveness also went up. In the case with the students possessing fewer speaking skills, however, such as those in freshman speaking classes in this study, this may hold true only under some conditions such as group work situations. The tendency for congruency between the time/frequency of activities and the scores of student evaluation is more evident in

sophomore intensive classes where students have more speaking skills.

It would be fair to say that there is a relationship between exposure to activities and the students' perception, although this may not be the case in individual work. One possible reason is that in order to have congruity between time/frequency of activities and students' sense of preference/effectiveness, students and teachers must share the learning objectives that are aimed to develop communication skills. Group work intends to enhance the chance of student interaction; a shared notion of communicative orientation between students and teachers may facilitate the agreement between time and frequency allotment of activities and preference/perceived effectiveness by students. The level of communication skills, however, may affect this. Students with fewer speaking skills may like less demanding tasks, and may be in need of scaffolding. This was not the case with intensive classes in which students had more experience in speaking English. Thus, they positively evaluated communicative activities, which consequently raise the scores of preference and effectiveness.

The results of analysis of beyond activity types indicated that, as time/frequency of activities increased, so did preference and effectiveness scores. There is a point of caution here, however, in interpreting this result because the number of observations was very limited; thus, this finding is still tentative.

Regarding the second research question with respect to COLT features, student evaluation scores reflected the results of the COLT observation regarding topic control and student modality. We found that in two sophomore intensive classes, students preferred group activities with more content control on their part, while in another class, students preferred moderate content control to free conversation. Even though this sounds contradictive, we are able to say that this does indicate that students do not prefer excessive student control. Another COLT feature in

favor of student evaluation was multimodal activities. This was positively evaluated in two sophomore intensive classes. On the other hand, students in another sophomore intensive class did not prefer single modality activities in which they read a longer text.

As a result of these analyses, we noticed that COLT Observation Scheme was a potentially useful observation tool in complimenting student evaluation, though there were points COLT Observation Scheme could clarify and those it could not. When we limited the domain of analysis, the relationship became clearer. For example, influential COLT features included topic control and student modality. However, a broadened domain of analysis made results less reliable.

We also recognized some drawbacks of the COLT Observation Scheme. This scheme may not be effective in classes in which students' initiatives are so strong that the instructor's interaction is on demand of the students while parallel and simultaneous group activities are employed. What is going on in each group is out of focus in this observation scheme, and no space is provided to record types of activities taking place in those group activities. By the same token, the COLT Observation Scheme does not function well in analyzing the sequence or the inner structure of activities, either. In addition, this observation scheme was developed in a teaching environment where target-language use as a means of communication is the norm. In a place like Japan, where students are inclined to interact in their first language, some assumptions of the COLT Observation Scheme may not hold true. According to the COLT Observation Scheme, the more interactions students have, the more communicative the classroom activities are. It does not presume the interaction is in the first language. Finally, the COLT Observation Scheme is intended to be used by researchers; thus, when its terms are used on a student evaluation form for class activities, students may not understand what the terms refer to.

This study includes the following limitations. The number of participants was small, especially in the intensive courses. Observations were based on one-shot visits. Also, the instructor's intentions for the activities were not confirmed. For further studies, it is recommended to investigate determinants of classification regarding types of communicative instruction. Also we suggest increasing the number of student participants, conducting a longitudinal study, and carrying out post hoc interviews of instructors.

Note

1 There are three ways students are organized in participating classroom activities. One is a teacher-centered activity in which the teacher interacts with the whole class and/or with individual students. Another is group work in which pairs or group of students work on an activity. The third is individual work in which students individually work on an activity.

References

Anno, Y. (2003). COLT wo mochiita kouritsu shougakkou eigokatstudou no jyugyou bunseki kenkyuu. [Research on class analysis of English activities in public elementary schools using COLT]. *CELES Bulletin, 3*, 185-192.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Erlbaum Associates.

Foster, P., Tonkyn, A., & Wigglesoworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375.

Fröhlich, M., Spada, N. & Allen, P. (1985). Differences in the communicative orientation of L2 classroom. *TESOL Quarterly, 19 (1)*, 27-57.

Panova, I & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36(4)*, 573-596.

Spada, N. & Fröhlich, M. (1995). *COLT observation scheme*. Sydney: National Center for English Language Teaching and Research, Macquarie University.

Sugita, Y. (2002). Communication wo shiya to shita eigo no jyugyou bunseki: jyugyou kansatsuhou (COLT) no jitsuyou kanousei [Classroom-centered research on English classes: A view of differences in communicative orientation]. *Bulletin of the Faculty of Education & Human Sciences, 3(2)*, 79-86.

Sugino, N., Koga, Y., & Kawashima, H. (2004). A longitudinal study of English language activities at a kindergarten: Focusing on shifts in instructional practices. *Annual Review of English Language Education in Japan, 15*, 71-80.

Yamada, Y. (1996). A case study of children's cross-cultural understanding in relation to the target-language use. *JASTEC Journal, 15*, 13-26.

Yoneyama, A., Oominato, Y., Suda, T., Maeno, H. & Yoshida, N. (2003). Gengo shuutoku ni motozuku jyugyou no kaizen – COLT wo mochiita jyugyoubunseki. [The improvement of a class observation scheme on language acquisition]. *Monbukagakusho itaku kenkyuu: Kyoushokukatei ni okeru kyouiku naiyou houhou no kaihatsu kenkyuu jigyou kenkyuu houkokusho*, 2001-2002 *(2)*.