

A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners

Natsuko Shintani

Language Teaching Research 2011 15: 137

DOI: 10.1177/1362168810388692

The online version of this article can be found at:

<http://ltr.sagepub.com/content/15/2/137>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language Teaching Research* can be found at:

Email Alerts: <http://ltr.sagepub.com/cgi/alerts>

Subscriptions: <http://ltr.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ltr.sagepub.com/content/15/2/137.refs.html>

A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners

Language Teaching Research
15(2) 137–158

© The Author(s) 2011

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1362168810388692

ltr.sagepub.com



Natsuko Shintani

University of Auckland, New Zealand

Abstract

The study reported in this article investigated the comparative effects of two types of treatment – one of which emphasized input and the other output – on the vocabulary acquisition of young EFL learners. In the input-based instruction, the students were not required to produce output whereas in the production-based instruction the students were required to produce output. Thirty-six Japanese children aged 6–8 were divided into three groups (input-based, production-based and control group), received six weeks instruction and took four types of vocabulary tests as a pre-, post- and delayed post-test. The findings provide further evidence that both input-based and production-based instruction lead to both receptive and productive vocabulary knowledge. In general, the results show similar levels of effects for input-based and production-based instruction on vocabulary acquisition. However, an examination of process features indicates that the input-based tasks provided opportunities for richer interaction for the learners than the production-based activities. This may explain the better performance of the input-based group on the task-based comprehension test and the same levels of achievement in the production tests despite relatively fewer opportunities for second language (L2) production.

Keywords

vocabulary acquisition, input-based instruction, production-based instruction, task-based language teaching, young learners

Corresponding author:

Natsuko Shintani, Department of Applied Language Studies and Linguistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
Email: n.shintani@auckland.ac.nz

I Introduction

Despite the long-running debate over the roles of input and output in second language acquisition and the numerous studies that have compared the effects of the two instructional options on the acquisition of grammatical features, little research has dealt with vocabulary acquisition. This is reflected in the majority of elementary EFL course books that have typically employed production-based instruction (i.e. involving the learners in speaking or writing). However, input-based instruction (i.e. focus on learners' comprehension of input in order to achieve the outcome) might be suitable for young learners if we assume that the known characteristics of first language (L1) vocabulary acquisition – such as fast-mapping or extensive comprehension knowledge prior to production knowledge – are applicable to the second language (L2) acquisition of young learners. The study reported in this article attempted to investigate the comparative effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. Literature relevant to the two options for teaching vocabulary is first reviewed before the details of the study itself are presented.

I Input-based vs. production-based instruction

Krashen's input hypothesis (1982, 1985, 1998) claimed that production serves only for generating comprehensible input from the interlocutor, and that output does not make a real contribution to acquisition. He based this claim on the fact that:

1. output, especially comprehensible output, is too scarce;
2. it is possible to attain high levels of linguistic competence without output; and
3. there is no direct evidence that output leads to language acquisition (Krashen, 1998).

Swain (1985), and Swain and Lapkin (1995), on the other hand, argued strongly that 'pushed output' leads the learner to notice L2 elements and modify their output, both of which contribute to acquisition. DeKeyser and Sokalski (1996, 2001) contrasted production vs. comprehension practice based on Anderson's (1993) skill acquisition theory, claiming that comprehension and production skills in an L2 can develop as a result of practice in the respective skills.

A number of studies have contrasted the effects of input and output on the acquisition of grammatical features (see Erlam, 2003). On the other hand, little research has explicitly compared the effects of input and output on vocabulary acquisition except for a few studies that have investigated reading and writing. Some studies, however, gauge the relative effects of oral input and output on L2 vocabulary acquisition by investigating different types of interaction (Ellis, 1993, 1995; Loschky, 1994; Ellis & Heimbach, 1997; Ellis et al., 1999; Nagata et al., 1999). The overall results indicate that although many of the studies show that interactionally modified input (i.e. input plus output) is better for acquisition (Ellis, 1993; Nagata et al., 1999), when the time on task is controlled, pre-modified input (i.e. input only) leads to acquisition of more words (Ellis, 1995). Studies also indicate the complex relationship between comprehension and acquisition (Ellis, 1995; Ellis & Heimbach, 1997). Two of

these studies (Ellis and He, 1999; de la Fuente, 2002), which included an output condition as one of the independent variables, will be considered in detail.

Ellis and He (1999) compared the effects of pre-modified input, interactionally modified input, and modified output on the comprehension of directions in a listen-and-do task and the acquisition of new words embedded in the directions. Participants were 50 students from six intermediate-level classes of the Intensive English Language Program at a university in the USA. They were divided into three groups with three different types of instruction: pre-modified input, interactionally modified input, and modified output. Participants were instructed to listen to or to give directions to place pieces of furniture in an apartment plan. Five post-tests were administered to them. Post-tests 1, 3, and 5 were comprehension tests involving a picture-matching task, which required the participants to match individual pictures with a list of target words. Post-tests 2 and 4 were production tests, in which the participants were asked to recall the labelled furniture after completing the tasks. The results show that the modified output group consistently achieved higher vocabulary scores in comprehension and production than either of the input groups. Ellis and He suggest two reasons for the superiority of the output condition:

- that producing new words helped learners to process them more deeply than simply hearing them; and
- that the modified output condition afforded the learners a qualitatively richer discourse experience because the students in the output condition perceived the task as a collaborative problem-solving activity while in the input condition the teacher treated the task as a kind of test.

In a later study, de la Fuente (2002) compared the effects of non-negotiated pre-modified input (NNPI), negotiation without 'pushed output' (Swain, 1985) (NIWO), and negotiation with pushed output (NIPO) on L2 learner's comprehension in a listen-and-do task and their acquisition of new words. Participants were 32 university students who were native speakers of English in five intact classes of intermediate Spanish in a basic language program. The participants engaged in oral interactions with native speakers (NSs) in pairs. In both the NNPI and NIWO groups, NSs (i.e. information suppliers) gave instructions to non-native speakers (NNSs) to locate and identify the 10 target words. Participants in the NNPI and NIWO groups repeated the task while participants in the NIPO group played the role of information supplier for the second task. Participants' vocabulary knowledge was measured in terms of receptive and productive skill. The results indicate that the NIPO group outperformed both of the input groups in the productive test but not in the receptive test. De la Fuente proposes three possible explanations for the superiority of the NIPO over the NIWO:

1. the NIPO led to a deeper level of processing that can be discussed in two stages; namely, 'decoding' and 'code breaking' (Cook, 1991);
2. the NIPO condition is more likely to result in form negotiation than the NIWO condition; and
3. pushed output would focus learners' attention on form, allowing them to process the word for production.

De la Fuente suggests that as no difference was found between the NIPO and the NIWO in the receptive test, the processing required in the NIPO condition may not be necessary in order to acquire the words receptively.

Those studies suggest the following regarding the effects of input and output on vocabulary acquisition:

1. Input-based tasks lead to both receptive and productive vocabulary acquisition.
2. Negotiated interaction (in these cases output conditions) seems to lead to better acquisition than pre-modified input alone.
3. Output can lead to deeper processing and direct learner's attention to form, resulting in better acquisition than input.

However, both Ellis and He, and also de la Fuente, investigated adolescent learners at non-beginner levels. One of the characteristics of young learners of L1 is 'fast-mapping', i.e. the phenomenon that very young children can learn a new word with very little exposure to the word (Clark, 1993). There is some evidence that fast-mapping does occur with pre-schoolers learning L2 (Ellis & Heimbach, 1997; Rohde & Tiefenthal, 2000). Also children learning their L1 have set up numerous categories of objects prior to the production of any words (Clark, 1993, p. 44). It might be possible, therefore, that young learners benefit from input-based instruction more than production-based instruction when acquiring L2 vocabulary compared to older learners. No attempt has been made, to my knowledge, to investigate the relative effects of input-based and production-based instruction on young L2 learners at beginner levels. The study reported here attempts to do this.

One possible operationalization of input-based instruction is input-based tasks. Ellis' (2003) definition of tasks is followed in the study reported below; tasks are activities with a primary focus on meaning, involving real-world processes of language use, with a clearly defined communicative outcome, and the learners use their own linguistic resources when performing the tasks. Tasks can be focused or unfocused. Focused tasks are tasks designed to induce learners' incidental attention to some specific linguistic form when processing either input or output (Ellis 2003). Ellis (2003, p.37) suggests that 'simple listening tasks can be devised that can be performed with zero competence in the L2 and that thus cater to the "silent period", which characterizes the early stages of acquisition for some learners.' As tasks should provide real-world use of language by stimulating learners to draw upon their own linguistic resources (Ellis, 2003, pp. 14–16), it is impractical to design tasks that exclusively require production for beginner learners. On the other hand, activities used in present-practice-produce (PPP) type of instruction typically focus on learners' production (see Ur, 1996), and a variety of such activities have been developed and widely used in the EFL classroom. The study reported here compares the effects of such production-based activities and input-based tasks on vocabulary learning.

2 Research questions

1. Does participation in input-based tasks enable young L2 learners to acquire new vocabulary?
2. Does participation in production-based activities enable young L2 learners to acquire new vocabulary?

Table 1 Number, age, and experience of the participants for each group

| Class (student number) | Ages | Experience |
|--------------------------------|------|------------|
| <i>IB group (n = 13):</i> | | |
| IB-A (n = 6) | 6–7 | 4 months |
| IB-B (n = 7) | 7–8 | 16 months |
| <i>PB group (n = 11):</i> | | |
| PB-A (n = 3) | 6–7 | 4 months |
| PB-B (n = 8) | 7–8 | 16 months |
| <i>Control group (n = 12):</i> | | |
| Control-A (n = 7) | 6–7 | 4 months |
| Control-B (n = 5) | 7–8 | 16 months |

3. Which instructional approach (input-based tasks vs. production-based activities) results in more newly acquired words?
4. To what extent do the instructional processes arising in the input-based and production-based interactions differ?

II Method

I Participants

Thirty-six Japanese children, aged 6–8, participated in this project. They were in six intact classes at a private English school in Japan. The participants received two 45-minute lessons per week and did not have any English instruction other than at school. The two experimental groups and one control group each consisted of two classes respectively: one class with 6–7-year-old students with four months of English learning experience, and one class with 7–8-year-old students with 16 months of experience, respectively. The number of the students in each class is shown in Table 1.

2 Design

The study employed a quasi-experimental pre-test–post-test design. The three groups received three different types of treatment:

1. the input-based (IB) group received input-based instruction;
2. the production-based (PB) group received production-based instruction; and
3. the control group received a set of three activities (English songs, Total Physical Response, and alphabet practice) without being exposed to any target words.

In the input-based instruction, the students were not required to produce output, whereas in the production-based instruction the students were required to produce output. However, the participants in both groups had opportunities for both input and output. Each group received six lessons (two lessons a week) during the project. As controlling time-on-task is one important factor for comparing different types of instruction (Ellis et al., 1994; Loschky, 1994, Ellis & He, 1999, de la Fuente, 2002), the lesson time for

Table 2 Overview of the research design

| | IB group | PB group | Control group |
|---------|------------------|-------------------------------------|-------------------|
| Week 1 | | Pre-test (2 production tests) | |
| Week 2 | | Pre-test (2 comprehension tests) | |
| Week 3 | Task lesson 1, 2 | Exercise lesson 1, 2 | Unrelated lessons |
| Week 4 | Task lesson 3, 4 | Exercise lesson 3, 4 | Unrelated lessons |
| Week 5 | Task lesson 5, 6 | Exercise lesson 5, 6 | Unrelated lessons |
| Week 6 | | Post-test 1 (2 production tests) | |
| Week 7 | | Post-test 1 (2 comprehension tests) | |
| Week 8 | | | |
| Week 9 | | | |
| Week 10 | | | |
| Week 11 | | Post-test 2 (2 production tests) | |
| Week 12 | | Post-test 2 (2 comprehension tests) | |

each group was set at approximately 30 minutes. All the lessons for the three groups were taught by the researcher. Four tests were designed to measure vocabulary knowledge, and were conducted over two weeks (i.e. two production tests in the first week and two comprehension tests in another). This testing regime was repeated three times as pre-test, post-test 1, and post-test 2. Pre-test was conducted two weeks prior to the first lesson. Post-test 1 was conducted one week after the treatment round, and post-test 2 was conducted four weeks after finishing post-test 1. All the treatment classrooms were video- and audio-recorded in order to investigate the instructional processes. The overview of the research design is shown in Table 2.

3 Target items

Twenty-four nouns were selected as target items, as nouns are considered easier for young learners to learn, being more likely to evoke images and are thus more meaningful than verbs or adjectives (Ellis & Beaton, 1993; Ellis, 1994). In order to select the target nouns, three familiar categories for Japanese children were chosen: animals, home appliances, and fruit and vegetables. Two criteria for selecting the items were used:

1. Japanese children had generally seen the real objects or pictures of the items, therefore the participants were assumed to be familiar with them; and
2. The items were not in the *List of English loanwords in Japanese* (Daulton 1999), avoiding the possibility of participants inferring the word meanings by using their L1 knowledge on the comprehension test.

Eight items were selected for each category as shown in Table 3.¹

The 24 items were introduced differently to the two experimental groups. In the PB group, the three sets of target items were introduced separately in each lesson: item A (8 items) for Lesson 1 and 2, item B (8 items) for Lesson 3 and 4, and item C (8 items) for

Table 3 Target vocabulary items

| | | |
|--------|---------|--|
| Item A | 8 items | camel, ostrich, crocodile, hippopotamus, squirrel, seal, polar bear, peacock |
| Item B | 8 items | pan, ladle, chopsticks, cutting board, plate, soap, battery, toothbrush |
| Item C | 8 items | green pepper, eggplant, chestnut, radish, leek, pear, mandarin, persimmon |

Lesson 5 and 6. In the IB group, on the other hand, all 24 target items were introduced each time in order to avoid the learners becoming aware of the linguistic goal by being exposed to just a small number of words many times in one lesson. Therefore, the same lesson consisting of the same activities was conducted six times with all the 24 items for the IB group but with different item groups for the PB group. In order to balance the time difference between exposures to the target items and post-tests, the learners in the PB group were given opportunities to review the previous target items in the last activity of each lesson. As there was no significant difference between the mean scores of the three item groups (8 items for each item group) attained by the PB group, it is assumed that separating the items for the PB group did not cause any imbalance in the learning of each item group.

4 Tasks used with the IB group

Three listen-and-do tasks were designed and conducted in one lesson that was repeated in each of the six lessons. The task goals (e.g. 'helping the zoo and the supermarket and collecting the cards') and task procedures were explained to the participants in Japanese at the beginning of each task. The teacher made every effort to use only English during the tasks with the help of non-verbal cues. Descriptions of each task are as follows:

Task 1: Help the zoo and the supermarket: Task 1 required the learners to listen to the teacher's commands and find the cards with the target items. Small flash-cards for each target item, one three-sided card holder, and one small empty box for incorrectly answered cards were placed in front of each participant. They were instructed to choose one of the cards while listening to the direction in English and then to hold it in their hand. All the participants displayed the card they had chosen at the same time followed by the teacher providing the correct answer to all the participants. Participants who had answered correctly placed the card in the holder, but the participants who had answered incorrectly replaced the card they had chosen wrongly on the table and then put the correct card in their 'incorrect' box. After finishing the commands for all items, each student counted the number of cards in his or her incorrect box. The student with the fewest cards was the winner.

Task 2: Help the animals: The same flash-cards as in Task 1 were used. Each student had a set of these cards. The participants were requested to find pairs of cards that corresponded to the teacher's statements. For example, if the teacher said 'The polar bear needs the battery', they had to find the card showing 'polar bear' and 'battery'. The students held up the pair cards they had selected so that the teacher could indicate whether they had chosen the correct or incorrect cards. They were allowed to change their selections until they were able to hold up the correct cards. This was repeated until all the animal cards had been used.

Task 3: Listening bingo game: This was a form of picture bingo. The participants were requested to choose nine cards out of the 10 cards they were provided with and to lay them out in a 3×3 formation in front of themselves. The teacher then called out a word naming one of the pictures. If a student had a card showing the picture of this word, he or she turned it face down. After six words had been presented, each participant's Bingo score (the number of lines of turned-over cards) was checked by the teacher. Two or three rounds were played depending on the available class time.

5 Activities used with the PB group

The activities for the PB group were also conducted by the researcher. One set of the five activities was conducted in each lesson and repeated for six lessons with different item groups. At the beginning of each lesson, the participants were informed of the purpose of the lesson (learning new vocabulary). The researcher used L2 (English) during the activities, but L1 (Japanese) was used to explain the methodology of activities when necessary. The teacher usually provided corrective feedback on the learners' production by way of recasts. Descriptions of each exercise are as follows:

Exercise 1: Listen and repeat: In this activity, flash-cards representing eight target items were presented one by one to the participants. The participants were requested to repeat the word after the teacher chorally. Each flash-card was presented twice. The activity was then repeated once.

Exercise 2: Guess the hidden items: In this activity, the teacher asked all the participants, in English, to name each flash-card while displaying and placing it face up in a holder one by one. When one card was left, the teacher placed the last card face down in the holder and asked the participants to try to name the hidden item. The teacher provided feedback, indicating whether a student was correct or incorrect, and recast wrong answers. This activity was repeated four times.

Exercise 3: Throwing dice: This activity also required the participants to verbalize the names of flash-cards. All the flash-cards were placed on the holder face down. Each card was numbered. One of the participants threw a dice. The teacher then picked up the card corresponding to the number shown on the dice and asked the participant to name the item. If the answer was wrong, the teacher provided a recast and turned the card face down again. If the answer was correct, the card was left face up. This was repeated until each participant had had two turns.

Exercise 4: Production bingo game: Each participant was given a set of flash-cards representing each target item, and instructed to lay them out in a 3×3 formation in front of themselves. The teacher also had a set of cards. The teacher asked one of the participants to pick up one card from his or her cards and to show it to the other participants. Then all the participants named it chorally and turned over the same card from their own set. This was continued for six cards. Each participant's bingo score (the number of lines of

Table 4 Types of learning measured by each test

| Type of input | Comprehension | Production |
|--------------------|--------------------------------|-------------------------------|
| Task-based test | Category task test | Same or Different task test |
| Discrete-item test | Multiple-choice listening test | Discrete-item production test |

turned-over cards) was checked by the teacher. Two or three rounds were played depending on the available class time.

Exercise 5: Kim's game: The six sets of flash-cards (two sets for each item groups) were spread randomly on the table face down. The participants took turns to turn over one of the cards, and then they all said aloud the item shown on the card. When a student turned a card over that was the same as one of the cards already face up, he or she was allowed to keep the pair of cards. If the participants could not name it, or provided a wrong answer, the teacher provided feedback using recasts. After all the cards had been turned over, the number of pairs collected by each participant was counted.

6 Recording and transcribing of lessons

In the lessons, students sat around a square table with eight seats. One video-camera was focused on the students, positioned at one side of the classroom where the teacher mostly stayed during the lesson. The audio-recorder microphone was attached to the teacher. All the audio-recorded data was transcribed.

7 Testing materials and procedures

Studies indicate the importance of using multiple assessments (Beretta & Davies, 1985; Schmitt, 1994; Laufer, 1998; Ellis, 2004, 2005; Takimoto, 2009). In the present study, four tests were designed for measuring learners' knowledge of target vocabulary items from two contrasted aspects (comprehension vs. production) and using two methodologies (task-based vs. discrete-item tests) as shown in Table 4.

The two comprehension tests were administered to the participants in groups. The two production tests were administered to participants individually by the researcher. In the production tests, ample time was given to answer, and when the participants provided incorrect answers three times consecutively, the test was stopped. Descriptions of each test are as follows.

a Multiple-choice listening test: This test required the participants to listen to an audio-recorded word and choose the appropriate picture from six pictures. The test contained 40 questions that included 24 target items together with 16 distractors. The 16 distractors, which had been selected from the *List of English loanwords in Japanese* (Daulton, 1999), were used to maintain participants' motivation to complete the test. Five seconds, which was predetermined to be enough for children aged 6–7 to answer, was given for

each question. The participants obtained one score for each item if they correctly chose the vocabulary item; the maximum score possible was 24.

b Category task test: A listen-and-do task was employed for this test. The participant was asked to listen to sentences and to decide in which of four given situations (fruit and vegetable shop, kitchen, bathroom, and zoo) each sentence had been said. The participants obtained one score for each item when they chose the correct category; the maximum score possible was 24.

c Discrete-item production test: In this test, the researcher asked the participant to name each flash-card for 24 target vocabulary items, which were displayed in turn. The participants obtained one point for each item correctly provided orally, irrespective of morphological errors. Participants' answers were considered to be correct when they provided the whole word even if there was a pronunciation error (e.g. *'toobrush'* for *'toothbrush'* or *'cuttin boar'* for *'cutting board'*), but it was considered to be incorrect when they failed to produce the whole word (e.g. *'bear'* for *'polar bear'* or *'chops'* for *'chopsticks'*).

d 'Same or different' task test: A two-way information-gap task called *'same or different?'* was employed for this test. The participant and the researcher had a different sheet showing 24 pictures of objects. The participant was asked to name each object, and the researcher then told the participant which picture she herself had. If the participant's and researcher's pictures were the same, they put a tick in the square showing the picture. If they were not the same they put a cross. The participants obtained one score for each item correctly provided orally. The criterion for correct production used for the discrete-item vocabulary production test was also employed for this test.

III Analysis

In order to answer research questions 1, 2 and 3, the total number of participants in each group, the mean scores, and the standard deviations for each test were calculated. Then the differences in the mean scores between the three repeated tests (research questions 1 and 2) and between the three groups (research questions 3) were analysed in a series of one-way analyses of variance (ANOVA) with post-hoc Bonferroni adjustments for multiple comparisons (alpha level = .05).

In order to answer research question 4, the transcriptions of the classroom interactions were analysed. The target word tokens in Lesson 3 of the IB and the PB group were counted to examine the students' L2 exposure and the opportunity for L2 production.² The tokens were further sorted in terms of those initiated by the teacher and the students, and in terms of whether they were used in an isolated manner or embedded in an utterance. The students' L2 utterances in Lesson 3 were also classified in terms of two types of interaction: IRF (i.e. initiate-respond-feedback) exchange (Sinclair & Coulthard, 1975) and negotiation of meaning (Long, 1983). The figures of class A and B were averaged and the difference between the IB and the PB groups were compared using chi-square test.

Table 5 Descriptive statistics for the four vocabulary tests

| | Pre-test | | Post-test 1 | | Post-test 2 | |
|--|----------|-----|-------------|-----|-------------|-----|
| | Mean | SD | Mean | SD | Mean | SD |
| <i>Multiple-choice listening test:</i> | | | | | | |
| IB group (n = 13) | 4.7 | 2.3 | 21.9 | 2.8 | 22.8 | 1.3 |
| PB group (n = 11) | 4.2 | 2.2 | 19.9 | 4.0 | 21.1 | 2.9 |
| Cntl. group (n = 12) | 4.2 | 2.1 | 4.7 | 2.1 | 4.8 | 1.6 |
| <i>Category task:</i> | | | | | | |
| IB group (n = 13) | 0.0 | 0.0 | 18.3 | 3.4 | 18.5 | 3.5 |
| PB group (n = 11) | 0.0 | 0.0 | 15.0 | 3.5 | 15.3 | 2.6 |
| Cntl. group (n = 12) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| <i>Discrete-item production test:</i> | | | | | | |
| IB group (n = 13) | 0.8 | 1.4 | 12.0 | 4.6 | 13.7 | 3.9 |
| PB group (n = 11) | 1.2 | 1.2 | 14.1 | 6.8 | 13.6 | 6.3 |
| Cntl. group (n = 12) | 1.1 | 1.2 | 1.3 | 1.1 | 1.8 | 0.8 |
| <i>Same or Different task:</i> | | | | | | |
| IB group (n = 13) | 0.0 | 0.0 | 8.5 | 2.7 | 10.5 | 3.4 |
| PB group (n = 11) | 0.0 | 0.0 | 11.1 | 5.6 | 11.1 | 6.1 |
| Cntl. group (n = 12) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

IV Results

Table 5 presents the descriptive data for the scores of the four vocabulary tests. The maximum obtainable score was 24 for each test. In order to answer research questions 1 and 2, the scores obtained by the three groups at Times 1, 2, and 3 were analysed using a series of repeated-measures ANOVAs (within-subjects effects) with post-hoc Bonferroni adjustment for multiple comparisons. For the multiple-choice listening test, a significant effect was found ($F(2, 66) = 579.615, p < .05$) and the following multiple comparisons test shows that the IB and the PB groups significantly improved on post-test 1 from the pre-test and on post-test 2 from the pre-test, but there was no significant difference between post-test 1 and post-test 2. There was no significant difference between the control group's three repeated tests. For the category task test, a significant effect was also shown ($F(2, 66) = 482.480, p < .05$). The multiple comparisons test shows that the IB and the PB groups performed significantly better on post-test 1 than on the pre-test and on post-test 2 than on the pre-test, but there was no significant difference between post-test 1 and post-test 2. There was no significant difference between the control group's three tests: the mean scores remained at nil throughout the repeated tests. As for the discrete-item production tests, a significant effect ($F(2, 66) = 137.747, p < .05$) was also found. The multiple comparisons test shows that the IB and the PB groups improved significantly on post-test 1 from the pre-test and on post-test 2 from the pre-test. Furthermore, the IB group performed significantly better on post-test 2 than post-test 1, while there was no significant difference between post-tests 1 and 2 for the PB group. There was no significant difference between the control group's three tests. The 'same or different' task also

Table 6 Summary of results for between tests differences

| Tests | IB group | | | PB group | | | Control group | | |
|--------------------------------|-------------|-------------|-------------|-------------|-------------|----------|---------------|-----------|----------|
| | Pre-Post1 | Pre-Post2 | Post 1-2 | Pre-Post1 | Pre-Post2 | Post 1-2 | Pre-Post1 | Pre-Post2 | Post 1-2 |
| Multiple-choice listening test | sig. | sig. | n.s. | sig. | sig. | n.s. | n.s. | n.s. | n.s. |
| Category task | sig. | sig. | n.s. | sig. | sig. | n.s. | n.s. | n.s. | n.s. |
| Discrete-item production | sig. | sig. | sig. | sig. | sig. | n.s. | n.s. | n.s. | n.s. |
| Same or Different task | sig. | sig. | sig. | sig. | sig. | n.s. | n.s. | n.s. | n.s. |

Notes: Pre-Post1: improvement from pre-test to post-test1; Pre-Post2: improvement from pre-test to post-test2; Post1-2: improvement from post-test1 to post-test2

had a significantly different effect ($F(2, 66) = 108.655, p < .05$). The multiple comparisons test showed that the IB and the PB groups achieved significantly higher scores in post-test 1 than in the pre-test and in post-test 2 than in the pre-test. Furthermore, the IB group performed significantly better on post-test 2 than post-test 1 while there was no significant difference between post-tests 1 and 2 for the PB group. The mean scores of the control group remained at nil throughout the three tests. The summary of the results are shown in Table 6.

In order to answer research question 3, repeated-measures ANOVAs (between-subjects effects) following the post-hoc Bonferroni adjustment for multiple comparisons were conducted to compare the scores of the three groups. For the multiple-choice listening test, a significant effect was found ($F(2, 33) = 122.032, p < .05$), and the post-hoc multiple comparisons test showed that both of the IB and the PB groups significantly outperformed the control group in post-tests 1 and 2, but there was no significant difference between the two experimental groups in post-tests 1 and 2. For the category task test, a significant effect was found ($F(2, 33) = 188.303, p < .05$). The post-hoc multiple comparisons test shows that both of the IB and the PB groups significantly outperformed the control group in post-tests 1 and 2, and also the IB group significantly outperformed the PB group in post-tests 1 and 2. As for the discrete-item production tests, a significant effect was found ($F(2, 33) = 162.900, p < .05$). The post-hoc multiple comparisons test showed that both of the IB and the PB groups significantly outperformed the control group in post-tests 1 and 2, but there was no significant difference between the two experimental groups in post-tests 1 and 2. For the 'same or different' task test, a significant effect was found ($F(2, 33) = 33.248, p < .05$). The post-hoc multiple comparisons test shows that both of the IB and the PB groups significantly outperformed the control group in post-tests 1 and 2, but there was no significant difference between the two

Table 7 Significant differences between the IB group and the PB group

| | |
|-----------------------------|---------------------|
| Same or Different task | IB group = PB group |
| Discrete-item production | IB group = PB group |
| Category task | IB group > PB group |
| Discrete-item comprehension | IB group = PB group |

Table 8 Teacher's and students' production of target words in Lesson 3

| | PB group | | IB group | |
|------------------------------|----------|---------|----------|---------|
| | Class A | Class B | Class A | Class B |
| Teacher's target word tokens | 235 | 202 | 323 | 328 |
| Students' target word tokens | 192 | 252 | 84 | 62 |
| Total target word tokens | 427 | 454 | 407 | 390 |

Table 9 Students' target word tokens initiated by teacher / students in Lesson 3

| | PB group | | IB group | |
|-------------------|----------|---------|----------|---------|
| | Class A | Class B | Class A | Class B |
| Teacher-initiated | 172 | 229 | 0 | 0 |
| Student-initiated | 20 | 23 | 84 | 62 |

experimental groups in post-tests 1 and 2. The summary of the test results for between group differences is shown in Table 7.

In order to answer research question 4, the classroom interaction transcription was analysed. First, the number of target word tokens produced by the teacher and the participants in one of the six lessons (Lesson 3) was counted. As Table 8 shows, the teacher provided 235 target word tokens in the PB-A class, 202 in the PB-B, 323 in the IB-A and 328 in the IB-B. The students provided 192 target word tokens in the PB-A, 252 in the PB-B, 84 in the IB-A and 62 in the IB-B. The target word tokens for each class total 427 in the PB-A class, 454 in the PB-B class, 407 in the IB-A class and 390 in the IB-B class. The results of chi-square test showed that there were significant differences between the two groups in the number of the teacher's tokens ($\chi^2 = 21.007$, $df = 1$, $p < .01$), and in the students' tokens ($\chi^2 = 75.258$, $df = 1$, $p < .01$). However, there was no significant difference in the total tokens between the two groups ($\chi^2 = 2.100$, $df = 1$, $p < .01$).

Students' target word tokens were further counted in terms of the words initiated by the teacher or by the students. 'Teacher-initiated' refers to target word tokens where the students repeated after the teacher, answered questions initiated by the teacher directly or indirectly. 'Student-initiated' represents target word tokens where the students produced private speech or social speech such as negotiation of meaning. As Table 9 shows, 172 teacher-initiated tokens occurred in the PB-A class, 229 in the PB-B, compared to no

Table 10 Number of isolated / embedded target word tokens by teacher in Lesson 3

| | PB group | | IB group | |
|----------|----------|---------|----------|---------|
| | Class A | Class B | Class A | Class B |
| Isolated | 235 | 183 | 98 | 57 |
| Embedded | 0 | 19 | 225 | 271 |

occurrence in either of the IB classes. As for student-initiated tokens, 20 tokens occurred in the PB-A class, 23 in the PB-B, 84 in the IB-A, and 62 in the IB-B. The results of chi-square test showed that there was a significant difference between the two groups in the number of the student-initiated tokens ($\chi^2 = 27.379$, $df = 1$, $p < .01$).

The number of target word tokens provided in an isolated or embedded manner in one of the six lessons was also counted. A typical occurrence of the isolated use is shown in the following example, which was often observed in the PB treatment, reflecting the characteristics of this type of instruction.

Teacher: What is it?
Students: ...
Teacher: Hippopotamus. Hippopotamus.

(PB-A: Lesson 1)

The following excerpt shows an example of target words embedded in sentences. This was mostly observed in the IB treatment.

Teacher: Next one. Please take the chopsticks to the supermarket. Chopsticks.
Student: Chopsticks?
Teacher: Yes, you use chopsticks every day. You use chopsticks when you eat rice.

(IB-A: Lesson 3)

As Table 10 shows, in the PB class the teacher provided 418 isolated target tokens and 19 embedded target tokens, whereas in the IB class the teacher provided 155 isolated tokens and 496 embedded tokens. The results of chi-square test showed that there were significant differences between the two groups in the number of the isolated tokens ($\chi^2 = 66.794$, $df = 1$, $p < .01$), and the embedded tokens ($\chi^2 = 2.196$, $df = 1$, $p < .01$).

In order to investigate the general characteristics of the discourse in each group, the number of IRF episodes and negotiation of meaning sequences that took place in Lesson 3 were also counted. IRF episodes occurred when one student or the whole class responded to the teacher's elicitation, as in this example:

Teacher: What's this?
Students: Plate.
Teacher: Good. Plate.

(PB-B Lesson 3)

Table 11 Number of students' IRF exchanges and negotiation sequences in Lesson 3

| | PB group | | IB group | |
|-------------|----------|---------|----------|---------|
| | Class A | Class B | Class A | Class B |
| IRF | 65 | 135 | 6 | 19 |
| Negotiation | 0 | 0 | 25 | 0 |

Negotiation of meaning takes place through the collaborative work that speakers undertake to achieve mutual understanding when there is some kind of communication problem (Ellis, 2008, p. 224). Negotiation sequences occurred when students attempted to negotiate the meaning of the target items, as in this example:

Teacher: Please take the mandarin to the supermarket.

Student: Fruit? Fruit?

(IB-A Lesson 3)

Negotiation also occurred when students used L2 in conjunction with gestures, as in this example:

Teacher: Turn over the radish.

Student: [Holding and showing his 'radish' card] Radish? Radish?

Teacher: Yes, yes.

(IB-A Lesson 5)

Table 11 shows the number of IRF episodes and negotiation sequences occurred in each class. Sixty-five utterances in IRF interactions occurred in the PB-A class, 135 in the PB-B, 6 in the IB-A and 19 in the IB-B. No negotiation occurred except in the IB-A class where there were 25. The results of chi-square test showed that there was a significant difference between the two groups in the number of IRF exchange episodes ($\chi^2 = 66.982$, $df = 1$, $p < .01$).

In summary, the teacher produced significantly more target words in the IB lessons whereas the students in the PB group produced more target words than those in the IB group. The total target word tokens were significantly more for the PB group than the IB group. The students in the IB group produced more student-initiated tokens than those in the PB group, while the teacher-initiated tokens predominantly occurred in the PB group. In the IB lessons, the teacher provided less isolated target words and more embedded target words than in the PB group. IRF exchanges occurred more often in the PB group than in the IB group while negotiation sequences occurred only in the IB group.

V Discussion

Research question 1 asked whether young L2 learners were able to acquire new vocabulary through input-based instruction. The answer to this question is yes. The IB group

made significant gains compared to the control group in post-tests 1 and 2 on all tests. The results are consistent with earlier findings that input-based instruction results in the acquisition of both receptive and productive knowledge (Ellis & He, 1999; de la Fuente, 2002).

Interestingly, the IB group significantly improved their scores on post-test 2 from post-test 1 on the production tests. If we look at the individual test scores, the IB participants demonstrated only receptive knowledge of some items in post-test 1, but successfully showed both receptive and productive knowledge of the same items in post-test 2. There were some items that were answered correctly on the multiple-choice listening test but not on the discrete item-production test in post-test 1, and then answered correctly on both the listening and production tests in post-test 2. The number of such items averaged 2.15 for the IB group and 0.82 for the PB group. In other words, the learners in the IB group first developed receptive knowledge and subsequently developed productive knowledge, whereas the PB group did not progress.

Research question 2 asked whether young L2 learners were able to acquire new vocabulary through production-based instruction. The answer to this question is also yes. The PB group made significant gains compared to the control group in post-tests 1 and 2 on all tests. The results also show that productive learning results in the acquisition of both receptive and productive knowledge (e.g., Ellis & Beaton, 1993). There were no significant improvements between any of the post-tests 1 and 2.³

Research question 3 asked which instructional approach (input-based tasks vs. production-based activities) resulted in more newly acquired words. The overall achievement of the two experimental groups was very similar; no significant difference emerged between the IB and the PB groups on three out of the four tests (i.e. the multiple-choice listening test, the discrete-item production test, and the 'same or different' task test). In one of the four tests (category task test), however, the IB group outperformed the PB group in both post-tests 1 and 2.

Research question 4 asked, to what extent the instructional processes arising in the input-based and production-based interactions differed. The analysis investigated both qualitative and quantitative features of classroom interaction. The target tokens show the quantitative feature of interaction, i.e. the amount of input and output that took place in each group. As Table 7 shows, the PB group produced significantly more target words than the IB group. However, if we assume that the participants were exposed to input from their peers as well as the teacher, the amount of input appears to be similar for each group; there is no significant difference in the total number of tokens between the two groups. In other words, although the learners in the PB group produced more target words, they had a similar amount of input to the IB group.

Three qualitative aspects were also analysed:

1. the degree of discourse control by the learners;
2. the occurrence of IRF exchanges and negotiation sequences; and
3. the manner in which the target words were provided.

The results show that the learners in the IB group experienced higher levels of discourse control than those in the PB group. The participants in the IB group exclusively produced target words initiated by the students, whereas in the PB group the students

produced words mainly initiated by the teacher (Table 6). Negotiation occurred only in one of the IB classes. The number of IRF exchanges in the two IB classes significantly differed from the PB classes, indicating interactional differences between the two groups (Table 7). The manner in which the target words were presented also appears to be different. In the PB group the teacher predominantly provided target words in an isolated manner whereas in the IB group the teacher provided target words in an integrated manner (Table 8).

How do these process features explain the results for research questions 1, 2 and 3? For research questions 1 and 2, it was found that both the input-based tasks and the production-based activities lead to both receptive and productive vocabulary knowledge. As Table 7 shows, the PB group experienced substantial input and output in the classroom. The IB group was also exposed to considerable input. Although the IB group produced fewer target words than the PB group, they still produced a substantial number (73 target tokens on average during a 30 minute lesson). It seems then that the amount of input and output that the two groups experienced was sufficient to improve their production of the target forms as well as their understanding of them.

For research question 3, although the category task test showed that the IB group significantly outperformed the PB group, the other three tests did not show any significant differences between the two groups despite the fact that IB and PB instruction are theoretically very different. Thus, three questions arise:

1. why the IB group performed better on the category task test;
2. why the PB group performed as well as the IB group on the discrete-item comprehension test; and
3. why the IB group performed as well as the PB group on the production tests.

In order to discuss the first question, the characteristics of the category task need to be considered. In the category task, the target words were provided in sentences in which the target words were embedded (e.g. 'Look. The peacock has beautiful feathers.'). The participants needed to parse the target words in sentences in order to accomplish the task goal. The opportunities for using parsing strategies in the two groups were not equivalent. As already mentioned, in the PB group the teacher predominantly provided target words in an isolated manner whereas in the IB group the teacher provided target words mainly in sentences. Studies have indicated that learners perform better on tests that are of the same type as the instruction in which they participated (Beretta & Davies, 1985). It might be possible, then, to assume that the situation in the IB classroom, where the participants needed to comprehend the embedded words more frequently than in the PB classroom, promoted the parsing strategies needed to identify the target words, thus leading to higher achievement on the category task test.

Concerning the second question, one possible explanation for the similar performance in the discrete-item comprehension test is that the level of explicitness of the instruction was similar. In the first activity for the PB group, the target items were presented orally by the teacher with the pictorial images. The tasks for the IB group also provided the learners with the opportunities to see the picture cards representing the target items and to hear the teacher naming the items, as the following example shows:

Teacher: Please take the seal to the zoo, seal, seal.

Student: *Wakatta* (I know).

Teacher: Ready? Three two one, go. Yes, (showing the picture card) this is a seal, seal, seal.

(IB-B: Lesson 5)

Studies suggest that the explicitness of instruction affects the level of vocabulary acquisition (Hulstijn, 1992; Penno et al., 2002). Thus, even though there was a distinction in the instructional approach, there may have been no difference from the learner's perspective. This may be the reason for the same levels of achievement of the two groups. Another possible reason for the similar performance is the similar amount of input. As the results show, the total number of target tokens for the two groups was not significantly different. Studies indicate that frequency of exposure is one of the major factors for vocabulary learning (e.g. Palmberg, 1987; Elley, 1989). The similar amount of exposure may be another reason for the similar achievement in the test.

However, why the IB group performed at the same level as the PB group on the production tests cannot be explained by the amount of production. The PB group produced significantly more target words than the IB group. The results do not support the claim of skill learning theory (Anderson, 1993), which claims that proceduralized knowledge involves different sets of skills for comprehension and production. It seems then, that the quality of the interactions rather than the quantity of the target words produced played a role, which I would like to discuss here.

The results show that the two groups were different in terms of the learners' degree of discourse control. The participants predominantly initiated production of the target words in the IB group while the teacher mainly initiated the production in the PB group. Various studies of first and second language learners (e.g. Wells & Montgomery, 1981; Ernst, 1994; Johnson, 1995; Ellis, 1999) have produced evidence that acquisition is more likely to be fostered when learners have the opportunity to exercise some degree of discourse control. It may be that such control promoted production knowledge of participants in the IB despite the fewer opportunities for production practice.

The two groups also differed in the occurrence of IRF exchanges and negotiation sequences. The results show that IRF interactions occurred mainly in the PB group, while negotiations occurred predominantly in the IB group. It has been shown that negotiation of meaning is profitable for comprehension, and may lead to greater acquisition than more restricted types of interaction such as IRF exchanges. Although the number of negotiated utterances was small (25 negotiated utterances out of 231 total utterances by students in the IB group), negotiation of meaning led to successful comprehension by the negotiator as well as by the other students, as Excerpt 1 shows. This student's negotiation (line 4 and line 8) led to this learner's successful comprehension as well as that of the other students in this class (line 14).

Excerpt 1 Comprehension of 'ladle' 2 (Lesson 2, 13:05 in IB-A)

1. *Teacher:* Please take the ladle to the supermarket.
2. *Student 2 and 3:* Ladle? Ladle?
3. *Teacher:* Yes. Can you find the ladle?

4. *Student 1:* Fruit?
5. *Teacher:* It's not fruit.
6. *Student 2:* Ladle?
7. *Teacher:* It's not fruit. Ladle ...
8. *Student 1:* *Miso* soup? *Miso* soup? (+) No?
9. *Teacher:* When you eat *miso* soup, maybe you need a ladle.
10. *Student 1:* Okay?
11. *Teacher:* Maybe, yes.
12. *Student 1:* Okay?
13. *Teacher:* Okay? (++) I don't know. Okay, ladle. (+) Three, two, one.
14. *All students:* [showing the correct card].
15. *Teacher:* Ladle. Everyone is correct.

This example of negotiation leading to successful comprehension is consistent with the findings of previous studies (Ellis et al., 1994; Ellis & Heimbach, 1997; Ellis & He, 1999; Van den Branden, 2000; de la Fuente, 2002). Although studies show a complex relationship between comprehension and acquisition, studies such as Ellis and He (1999) and de la Fuente (2002) indicate that the occurrence of negotiation leads to better vocabulary acquisition. This excerpt indicates that the input-based tasks in this study provided the context that induced learners to negotiate meaning. This might have resulted in the development of production knowledge of the target vocabulary in the input-based group despite the fewer opportunities for production practice.

The results also indicate that young EFL learners aged 6–8 can negotiate meaning using L2. Previous studies have reported children's ability to negotiate meaning (Ellis & Heimbach, 1997; Van den Branden, 1997, 2000; Oliver, 1998, 2002), although most of these have investigated children in mid or late primary school, with the exception of Ellis and Heimbach (1997) who reported that some children aged 5–6 had difficulty with negotiation of meaning in their L2. The present study indicates that 6–8-year-old children with very limited L2 skill can negotiate meaning. It should be noted, however, that only two students out of the 13 in the group were responsible for most of the negotiation. The others may have been in the process of acquiring negotiation skills.

VI Conclusions

This study has investigated the comparative effects of input-based and production-based instruction on vocabulary acquisition. Despite the fact that IB and PB instruction is theoretically different, there were also some similarities in terms of implementation. The results show that the input-based instruction provided opportunities for learners to produce, and the production-based instruction also provided opportunities for learners to comprehend input. Both experimental treatments led to gains in both comprehension and production vocabulary knowledge. The similar performance of the two groups in the comprehension test might be explained by the amount of input and explicitness of instruction. However, there were also a number of differences in interaction. The IB instruction provided certain kinds of opportunities that seem to foster development in a way that the PB did not. This might explain the IB group's better performance in the category task, the long-term effects on production, and the same levels of achievement

in the production tests despite fewer opportunities for production practice. The process data strongly suggests that the IB condition provided richer opportunities for learning than the PB group.

There are certain limitations in this study that need to be pointed out. First, the number of participants was relatively low, which limits generalization of these results. Another limitation is the representativeness of the participants, which also affects the generalizability of the results. The participants in the present study were students in a private English language school and they may have differed in the outcomes of their tests and their responses to the instruction from the general population of the same age group. As compulsory English education is due to be implemented in primary school in Japan, conducting the same experiment with learners of the same age with no experience of English would strengthen the generalizability of the results.

Despite the shortcomings mentioned above, the present study contributes to our understanding of the relative effects of input-based and production-based instruction on vocabulary acquisition for young EFL learners. As for the pedagogical implications, this study provides evidence that input-based tasks can be successfully implemented in EFL classrooms for young beginners and are at least as effective as production activities where vocabulary learning is concerned.

Notes

- 1 For the purposes of another study some of the items were provided in both singular and plural form. But for the purpose of this study the acquisition of the lexis was measured irrespective of whether or not the participants used plural or singular forms correctly.
- 2 Lesson 3 was chosen for qualitative analysis because it was in the middle of the six lessons. Learners spoke very little in Lessons 1 and 2. Also, due to technical problems with recording the lessons, only the data for Lesson 3 were complete.
- 3 There is a possibility of a ceiling effect in the multiple-choice listening post-test 1, as nearly all students scored nearly 100%. However, it is noteworthy that the other measures did not show any significant change from post-test 1 to post-test 2 despite there being room for progress.

References

- Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Beretta, A. & Davies, A. (1985). Evaluation of the Bangalore Project. *ELT Journal*, 39, 121–27.
- Clark, E.V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Cook, V. (1991). Review: Universal grammar and second language acquisition. *System*, 19, 342–46.
- Daulton, F.E. (1999). List of high-frequency baseword vocabulary for Japanese EFL students. *The Internet TESL Journal*, 7.
- DeKeyser, R.M. & Sokalski, K.J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46, 613–42.
- DeKeyser, R.M. & Sokalski, K.J. (2001). The differential role of comprehension and production practice. *Language Learning*, 51(s1), 81–112.
- de la Fuente, M.J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24, 81–112.
- Elley, W.B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24, 174–87.

- Ellis, N. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit and explicit learning of language* (pp. 213–82). London: Academic Press.
- Ellis, N. & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A*, 46, 533–58.
- Ellis, R. (1993). Naturally simplified input, comprehension and second language acquisition. In M.L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 53–68). Singapore: SEAMEO Regional Language Centre.
- Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16, 409.
- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam/Philadelphia: John Benjamins.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54, 227–75.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*, 27, 141–72.
- Ellis, R. (2008). *The study of second language acquisition*. Second edition. Oxford/New York: Oxford University Press.
- Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21, 285–301.
- Ellis, R. & Heimbach, R. (1997). Bugs and birds: Children's acquisition of second language vocabulary through interaction. *System*, 25, 247–59.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44, 449–91.
- Ellis, R., Heimbach, R., Tanaka, Y., & Yamazaki, A. (1999). Modified input and the acquisition of word meanings by children and adults. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 63–114). Amsterdam/Philadelphia: John Benjamins.
- Erlam, R. (2003). Evaluating the relative effectiveness of structured-input and output-based instruction in foreign language learning. *Studies in Second Language Acquisition*, 25, 559–82.
- Ernst, G. (1994). Talking circle: Conversation and negotiation in the ESL classroom. *TESOL Quarterly*, 28, 293–322.
- Hulstijn, J.H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P.J.L. Arnaud & H. B'ejoint. (Eds.), *Vocabulary and applied linguistics* (pp. 113–25). London: Macmillan.
- Johnson, K. (1995). *Understanding communication in second language classroom*. Cambridge: Cambridge University Press.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. London/New York: Longman.
- Krashen, S. (1998). Comprehensible output? *System*, 26, 175–82.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19, 255–71.
- Long, M.H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, 17, 359–82.
- Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition*, 16, 303–23.
- Nagata, H., Aline, D., & Ellis, R. (1999). Modified Input, language aptitude and the acquisition of word meanings. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 133–49). Amsterdam/Philadelphia: John Benjamins.

- Oliver, R. (1998). Negotiation of meaning in child interactions. *The Modern Language Journal*, 82, 372–86.
- Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *The Modern Language Journal*, 86, 97–111.
- Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners. *Studies in Second Language Acquisition*, 9, 201–19.
- Penno, J., Wilkinson, I., & Moore, D. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, 94, 23–33.
- Rohde, A. & Tiefenthal, C. (2000). Fast mapping in early L2 lexical acquisition. *Studia Linguistica*, 54, 167–74.
- Schmitt, N. (1994). Vocabulary testing: Questions for test development with six examples of tests of vocabulary size and depth. *Thai TESOL Bulletin*, 6, 9–16.
- Sinclair, J.M. & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–56). Rowley, MA: Newbury House.
- Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371–91.
- Takimoto, M. (2009). The effects of input-based tasks on the development of learners' pragmatic proficiency. *Applied Linguistics*, 30, 1–25.
- Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- Van den Branden, K. (1997). Effects of negotiation on language learners' output. *Language Learning*, 47, 589–636.
- Van den Branden, K. (2000). Does negotiation of meaning promote reading comprehension? A study of multilingual primary school classes. *Reading Research Quarterly*, 35, 426–43.
- Wells, G. & Montgomery, M. (1981). Adult-child interaction at home and at school. In P. French & M. McLure (Eds.), *Adult-child conversation* (pp. 210–43). New York: St Martin's Press.