

## **The Effects of a 24-Session EFL Pronunciation Course as Reflected in Learners' Self-Reports**

**SHIZUKA, Tetsuhito**

**Kansai University**

### **Abstract**

The present study examined the effects of a 24-session pronunciation course on participants' beliefs and perceived abilities, through their responses to survey questions. Japanese junior high school students ( $n = 66$ ) participated in a pronunciation course, the main activity of which was one-on-one performance coaching/testing. A Likert-scale survey of respondents' perceived abilities regarding segmental and suprasegmental features were continually conducted throughout the course. The same survey was also administered to a control group, learning English at the same school but not participating in the course. Sign tests indicated that only in the case of the treatment group did the perceived pronunciation abilities significantly and substantially increase after a year. Rasch analyses of participants' responses indicated that their pre- and post-course perceived abilities were markedly different, and that difficulties of different segmental and suprasegmental elements relative to each other were rather stable over time. Qualitative and quantitative analyses of verbal comments collected at the final session revealed that participants were generally highly satisfied with the course mainly because they (a) felt their abilities were improved, (b) enjoyed the one-on-one performance testing, and (c) were convinced of the importance of proper pronunciation.

**Key Words:** pronunciation, one-on-one performance testing, perceived abilities, Rasch analyses

### **1. Introduction**

The status of no other component in EFL teaching has rollercoasted so dramatically like that of pronunciation has. For the past 50 or more years, various methods and approaches placed this skill either at the forefront of instruction or in the back wings (Celce-Murcia, Brinton, & Goodwin, 1996; Morely, 1991; Levis, 2005). Today, with the prevalence of Communicative Language Teaching with its focus on language as communication, there finally seems to be a consensus among the teaching profession that good-enough pronunciation is "an essential component of communicative competence" (Morely, 1991, p.488).

How good is good enough, though, is a tricky question, particularly in the age of world Englishes (Smith, 1983). With non-native speakers of English in the outer and the expanding circles (Kachru, 1985) outnumbering native speakers by the ratio of approximately 3 to 1 (Crystal, 2003), Jenkins (2000) has proposed the Lingua Franca Core (LFC) as a more

practical and suitable goal for most learners of English. Promising as the LFC may appear, there is yet to be an agreement as to its desirability as a new pronunciation model (Dauer, 2005). In addition, even when the LFC is adopted, there is little doubt that typical Japanese learners of English need to acquire much better pronunciation because virtually all the phonological features of the English language at which Japanese learners are known to be weak are retained in Jenkins' pedagogical core.

Generally, Japanese learners of English are strongly influenced by their native language at segmental, suprasegmental (Flege, 1980; 1981) and voice quality setting (Esling & Wong, 1983) levels. They are poor at crucial phoneme distinctions (Jenkins, 2000; Arimoto, 2005), particularly at differentiating /r/ and /l/ (Flege, Takagi, & Mann, 1995; Riney, Takagi, & Inutsuka, 2005), as well as at using a wide-enough pitch range (Todaka, as cited in Celce-Murcia, Brinton, & Goodwin, 1996). This often results in pronunciation that breaks down communication with other non-native speakers (Jenkins, 2000), and which is, compared with native-like pronunciation, less comprehensible as well as less attractive, even to those who share their native language (Nakanishi, 2004). It is not surprising that an early study by Suter (1976) found Japanese speakers of English much less intelligible than their Arabic, Persian, and Thai counterparts.

Clearly, then, something needs to be done. Unlike dance steps, sounds of a new language cannot be learned simply by watching the instructor's movements: unlike a dancer's arms and legs, a language teacher's articulatory organs are mostly not in view (Yamada, Adachi, & ATR Institute, 1999). For that reason, as Arimoto (2005) rightly points out, the teacher's role as a coach is particularly important in pronunciation teaching. However, empirical studies on the effects of teaching pronunciation to Japanese students are rare. One of the few reports that the author has come across, by Asami and Tanaka (2005), is rather sketchy, which makes it difficult to interpret their results. Researchers and practitioners concur (Makino, 2005; Asami & Tanaka, 2005; Kosuge, 2005) that a major challenge in pronunciation teaching lies in maintaining learners' motivation until their knowledge is proceduralized, rather than in imparting declarative knowledge of articulatory phonetics. Unfortunately, proposals of systematic ways for enhancing and maintaining students' motivation towards pronunciation are again hard to find.

## **2. One-on-One Testing as a Motivator**

Possibly, the only known attempt in that direction is the personal card method (PCM) reported by Shizuka (1995). In the PCM, each student receives a personal card that specifies an appropriate number of pronunciation items—phrases or short sentences including target sounds—and they are told to train themselves outside class to master them. The students are encouraged to approach the teacher whenever and wherever possible out of class and to try orally producing any of the items, before the due date, which is, say, a week away. When a student's performance on an item is judged to be satisfactory, the student earns a point for that item. The total points earned before the due date will be the student's pronunciation score, which will account for a certain percentage in the final term grade. Hence, the essence of the PCM is cyclical one-on-one procedural-knowledge testing, as opposed to one-shot collective declarative-knowledge testing. Shizuka (1995) found that using this method three

times over a nine-week period was effective in significantly improving the accuracy rate of college students' /r/ sounds from 44 to 80 percent, as well as in enhancing their self-reported motivation towards bettering pronunciation.

The current study investigates the effects of a program that used what could be referred to as a more intensive version of Shizuka's PCM. In the academic year 2007, the author taught a 24-session course in English pronunciation for junior high 3rd year students. The aim of the program was to improve participants' skills in key segmental and suprasegmental features of English phonology. Two attempts were made to measure the effects of the course. One was administering a Likert scale survey (see below) continually to track possible changes in participants' self-perceived abilities and motivations. The other was audio-recording students' pronunciation of the same formulaic expression before and after the 24-week course, to submit the recordings to auditory and acoustic analyses. The purpose of the present paper is to report on the findings from the survey results. Those from the auditory and acoustic analyses of the recordings will be reported elsewhere (Shizuka, 2008).

### 3. The Study

The present case study examined whether or not the 24-session EFL pronunciation course based on the PCM was effective in affecting participants'

- (1) perceived importance of pronunciation;
- (2) level of attention paid to pronunciation when speaking English outside the course;
- (3) perceived skills in pronouncing target segmental sounds (/r/, /f/, /v/, /θ/, /ð/);
- (4) perceived skills in the target suprasegmental feature (stress-timed rhythm);
- (5) attitudes to and beliefs about anything else related to English.

#### 3.1 Participants

Participants of the course were 66 (31 males and 35 females) 3rd-year students (age 14–15) at a private junior high school in western Japan. These students signed up for the elective course on a forced-choice basis; all the students in the 3rd year were required to choose one of the five elective courses offered in different subject domains. The junior high school was attached to a high school, which in turn was attached to a university. This ensured that the 3rd-year students were generally free from the common pressure to cram for entrance examinations to high schools. The 66 students were divided into two groups (32 and 34) to participate in the course consecutively on the same days. The first group was trained in the first two 50-minute periods and the second group in the second two 50-minute periods on the same mornings. The course contents for the two groups were identical. The 66 participants as a whole will be henceforth referred to as the treatment group. Students in the 3rd year at the school who took other elective courses served as controls ( $n = 190$ ). No student in the treatment or the control group had lived in an English-speaking country longer than a month. The actual numbers of cases differed from analysis to analysis, due to missing values.

#### 3.2 Course Details

The course consisted of 24 sessions, 10 in Term 1, 9 in Term 2, and 5 in Term 3. One

session lasted for 100 minutes, with a 10-minute break in the middle. Course materials, which were 21 Handouts and 21 Personal Cards, were written by the author. A Handout presented short utterances from a movie clip and/or selected lines from a pop song that contained target phonological features of the day. A Personal Card listed seven or eight “items,” which were utterances or song lines selected from the Handout. Main segmental targets were consonants known to pose difficulties to Japanese learners (e.g., /r/, /f/, /v/, /θ/, /ð/), though a few vowels (/æ/, /ə:/) were focused on as well. The first 21 sessions had only segmental features as explicit testing points, treating suprasegmental features like word stress, linking, intonation only incidentally, while the last three included stress-timed rhythm among focal points.

A typical structure of one session was as follows: the first 15 to 20 minutes were spent on teacher-fronted explanations and after-the-model collective repetitions, using the Handout for the day. In this phase, relevant parts of movie clips or pop songs were often presented as models. The remaining 70 to 75 minutes were spent on cyclical one-on-one performance testing/coaching based on the Personal Card for the day. The 30-plus students formed a circle with their Cards in hand, and the author walked around inside the circle, testing on a one-on-one basis. A student pronounced one item without looking at the Card, and if the targeted sounds were all produced satisfactorily, the author declared a “pass.” When one or more features were not acceptable, the author declared a “fail” and quickly pointed out what was wrong (e.g., “Work on the /r/ in *very*”). (The author has an MA in TESOL, a PhD in Applied Linguistics, and ample experience in training secondary school teachers in pronunciation.) One who earned a “pass” drew a small circle in the designated box on the Card, while one who failed jotted down the reason for the failure in the box. A student was allowed to try the items only one at a time, whether resulting in a pass or a fail. One who failed in an item was expected to try the same item when subsequent turns came around, until a pass was earned for that item. The number of turns that came around for one student in one session was somewhere between 15 and 30. Near the end of a session, the testing rounds were declared to be over and the Cards were collected. The number of “passes” or circles the student earned was his or her score for the day. The cumulative scores of all the classmates were displayed, from time to time, in the form of a bar chart for everyone to inspect, to nurture friendly competition.

### 3.3 Survey

The survey asked the respondent to rate the degrees of their agreements with the 15 statements shown in Table 1 (originally in Japanese), on a 5-point Likert scale, from 4 (Absolutely true) to 0 (Not at all true).

**Table 1 Statements Used in the Survey, Each Preceded by Its Abbreviation**

1. [Imptnt]: Acquiring good pronunciation is important when learning English.
2. [Attntn]: When reading English aloud, I pay attention to my pronunciation.
3. [Cnfdnt]: I am fairly good at English pronunciation.
4. [R\_wA]: When I am paying attention, I can pronounce the “r” sound (as in right, America) correctly.
5. [R\_w/oA]: Even when I am not paying attention, I can pronounce the “r” sound (as in right, America) correctly.
6. [TH\_wA]: When I am paying attention, I can pronounce the voiceless “th” sound (as in think, thank) correctly.
7. [TH\_w/oA]: Even when I am not paying attention, I can pronounce the voiceless “th” sound (as in think, thank) correctly.
8. [DH\_wA]: When I am paying attention, I can pronounce the voiced “th” sound (as in the, other) correctly.
9. [DH\_w/oA]: Careless Even when I am not paying attention, I can pronounce the voiced “th” sound (as in the, other) correctly.
10. [F\_wA]: When I am paying attention, I can pronounce the “f” sound (as in fine, office) correctly.
11. [F\_w/oA]: Even when I am not paying attention, I can pronounce the “f” sound (as in fine, office) correctly.
12. [V\_wA]: When I am paying attention, I can pronounce the “v” sound (as in village, have) correctly.
13. [V\_w/oA]: Even when I am not paying attention, I can pronounce the “v” sound (as in village, have) correctly.
14. [RTM\_wA] When I am paying attention, I can read aloud English with proper rhythm.
15. [RTM\_w/oA] Even when I am not paying attention, I can read aloud English with proper rhythm.

*Note:* “Imptnt” stands for “Important”; “Attntn” for “Attention”; “Cnfdnt” for “Confident”; “wA” for “with attention”; “w/oA” for “without attention”; TH for /θ/; and DH for /ð/; These abbreviations will be used in Table 4.

The first three items tapped into pronunciation-related overall attitudes and confidence, while the remaining 12 concerned more specific skills of producing five target consonants and stress-timed rhythm, when attention was and was not being paid, separately. One interest was to find out whether or not Japanese EFL junior high school students’ self-reports will reflect the well-documented phenomenon of style-shifting (Labov, 1966; 1969; Dickerson, 1975; Schmidt, 1977; Sato, 1985; Shizuka, 1993), where a sound correctly produced when attention is being paid to pronunciation can be mispronounced when attention is paid to something else.

### 3.4 Procedure

With regard to the treatment group, the survey was conducted during the program sessions. Although the author’s first intention was to give the survey at the end of every session all through the year, for class-management and other reasons, the number of times it was done was 12 in all: six times in Term 1, once in Term 2, and five times in Term 3.

The control group answered the survey during English classes in which the focus was oral communication skills in general. This was done with the help of two English teachers who were in charge of those classes, twice: first in April, 2007 and second in February 2008. That is, the time when the treatment group answered the survey for the first time and that when the control group did so were approximately the same (both near the beginning of the academic year); the time when the treatment group filled out the survey for the last time and that when the control group did so for the second time were close to each other (both near the end of the academic year).

In the final session, students in the treatment group were also asked (a) to choose the degree to which they felt their pronunciation had improved on a 6-point scale (5 “very much”, 0 “not at all”), (b) to assign a percentage value to the degree of their overall satisfaction with the course, and (c) to verbally comment on the course. In every case, the students were instructed to provide their names.

Figure 1 illustrates the chronological relationships among the 14 survey administrations, 12 to the treatment group and 2 to the control group. Hereafter, for convenience, the survey administration to the treatment group at Time *X* will be referred to as TTX, and that to the control group at Time *Y* as CTY.

|                               |      |      |      |      |      |      |      |      |      |      |      |      |      |
|-------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Treatment Grp ( <i>n</i> =64) | TT01 | TT02 | TT03 | TT04 | TT05 | TT06 | TT07 | TT08 | TT09 | TT10 | TT11 | TT12 |      |
| Control Grp ( <i>n</i> =190)  | CT01 |      |      |      |      |      |      |      |      |      |      |      | CT02 |

**Figure 1 Timeline of Survey Administration.**

## 4. Results

### 4.1 Change in Mean Ratings

First, the mean of the ratings for each item in each administration was calculated (Table 2). Even though the author is fully aware that ordinal responses to Likert scale items do not lend themselves very well to mathematical manipulations (Nunnally & Bernstein, 1996, p. 16), this was done to quickly grasp the overall picture. A brief inspection of the values indicates that generally the ratings by the treatment group increased almost monotonically overtime from TT01 to TT13, whereas those by the control group rarely showed clear differences between CT01 and CT02.

**Table 2 The Mean Rating in Each Item in Each Survey Administration**

|        | TT01 | TT02 | TT03 | TT04 | TT05 | TT06 | TT08 | TT09 | TT10 | TT11 | TT12 | TT13 | CT01 | CT02 |
|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Item01 | 3.48 | 3.56 | 3.68 | 3.64 | 3.56 | 3.73 | 3.71 | 3.74 | 3.63 | 3.68 | 3.69 | 3.83 | 3.42 | 3.31 |
| Item02 | 2.63 | 2.94 | 3.18 | 3.26 | 3.13 | 3.32 | 3.35 | 3.37 | 3.24 | 3.30 | 3.21 | 3.38 | 2.57 | 2.65 |
| Item03 | 1.03 | 1.30 | 1.80 | 1.91 | 2.00 | 1.92 | 2.08 | 2.08 | 2.08 | 2.00 | 2.16 | 2.28 | 1.30 | 1.39 |
| Item04 | 1.67 | 1.98 | 2.43 | 2.73 | 2.64 | 2.94 | 2.82 | 2.95 | 2.95 | 2.86 | 2.79 | 3.14 | 1.99 | 2.21 |
| Item05 | 0.89 | 1.25 | 1.58 | 1.76 | 1.78 | 2.09 | 2.02 | 2.20 | 2.19 | 2.13 | 2.11 | 2.25 | 1.18 | 1.31 |
| Item06 | 1.77 | 2.30 | 2.95 | 3.02 | 2.91 | 3.12 | 3.00 | 3.03 | 2.97 | 2.98 | 3.02 | 3.25 | 2.12 | 2.13 |
| Item07 | 1.03 | 1.38 | 1.82 | 1.95 | 1.95 | 2.29 | 2.12 | 2.28 | 2.22 | 2.19 | 2.18 | 2.42 | 1.47 | 1.31 |
| Item08 | 1.72 | 2.08 | 2.89 | 2.80 | 2.77 | 3.03 | 2.89 | 2.97 | 2.85 | 2.92 | 2.93 | 3.20 | 1.96 | 2.00 |
| Item09 | 1.00 | 1.30 | 1.82 | 1.83 | 2.17 | 2.15 | 2.14 | 2.20 | 2.08 | 2.21 | 2.11 | 2.38 | 1.42 | 1.27 |
| Item10 | 1.94 | 2.17 | 3.00 | 2.89 | 2.84 | 2.98 | 3.12 | 3.17 | 3.07 | 3.00 | 2.98 | 3.26 | 2.21 | 2.22 |
| Item11 | 1.20 | 1.30 | 1.85 | 1.97 | 1.97 | 2.18 | 2.32 | 2.32 | 2.22 | 2.25 | 2.30 | 2.45 | 1.58 | 1.54 |
| Item12 | 1.92 | 2.24 | 3.03 | 2.82 | 2.92 | 2.97 | 3.00 | 3.02 | 3.03 | 2.97 | 2.87 | 3.18 | 2.06 | 2.16 |
| Item13 | 1.17 | 1.41 | 1.89 | 1.86 | 2.02 | 2.24 | 2.21 | 2.22 | 2.25 | 2.30 | 2.16 | 2.38 | 1.40 | 1.45 |
| Item14 | 1.19 | 1.87 | 2.46 | 2.44 | 2.39 | 2.56 | 2.45 | 2.51 | 2.41 | 2.52 | 2.38 | 2.60 | 1.91 | 1.79 |
| Item15 | 0.72 | 1.24 | 1.57 | 1.62 | 1.69 | 1.85 | 1.73 | 1.82 | 1.80 | 1.84 | 1.87 | 2.11 | 1.23 | 0.99 |

## 4.2 Sign Test on Each Item

Next, a sign test, a non-parametric equivalent of a paired  $t$  test, was conducted regarding TT01-TT12, and CT01-CT02, median differences, separately (Table 3). The results indicate that in the case of the treatment group, the responses collected at the beginning of and at the end of the course (TT01 and TT12) indeed had different medians for every item at  $p < .000$  level; the median ratings were significantly higher at the end. The effect size (Cohen's  $d$ ) was medium for Item 01 and large for all the other items. On the other hand, the ratings by the control group did not change significantly, except for Item 04 (R\_wA), which showed a significant increase ( $p = .034$ ) of small effect size ( $d = -.26$ ), and Item 15 (Rtm\_w/oA), which showed a significant *decrease* ( $p = .002$ ) of medium size ( $d = -.39$ )

**Table 3 Sign Test Results on TT01-TT12 and CT01-CT02 Differences, for Each Item**

| Item | TT01 vs. TT12 |      |       | CT01 vs. CT02 |      |       | Item | TT01 vs. TT12 |      |       | CT01 vs. CT02 |      |       |
|------|---------------|------|-------|---------------|------|-------|------|---------------|------|-------|---------------|------|-------|
|      | $Z$           | $p$  | $d$   | $Z$           | $p$  | $d$   |      | $Z$           | $p$  | $d$   | $Z$           | $p$  | $d$   |
| #01  | -3.59         | 0.00 | -0.44 | -1.12         | 0.26 | -0.14 | #09  | -6.57         | 0.00 | -0.81 | -1.30         | 0.20 | -0.16 |
| #02  | -5.53         | 0.00 | -0.68 | -0.57         | 0.57 | -0.07 | #10  | -6.57         | 0.00 | -0.81 | -0.32         | 0.75 | -0.04 |
| #03  | -6.36         | 0.00 | -0.78 | -1.54         | 0.12 | -0.19 | #11  | -5.69         | 0.00 | -0.70 | -0.42         | 0.67 | -0.05 |
| #04  | -6.94         | 0.00 | -0.85 | -2.12         | 0.03 | -0.26 | #12  | -6.52         | 0.00 | -0.80 | -0.90         | 0.37 | -0.11 |
| #05  | -6.44         | 0.00 | -0.79 | -1.69         | 0.09 | -0.21 | #13  | -5.69         | 0.00 | -0.70 | -0.68         | 0.50 | -0.08 |
| #06  | -7.28         | 0.00 | -0.90 | -0.45         | 0.65 | -0.06 | #14  | -6.62         | 0.00 | -0.82 | -1.22         | 0.22 | -0.15 |
| #07  | -6.52         | 0.00 | -0.80 | -1.87         | 0.06 | -0.23 | #15  | -6.93         | 0.00 | -0.85 | -3.14         | 0.00 | -0.39 |
| #08  | -7.07         | 0.00 | -0.87 | -0.68         | 0.50 | -0.08 |      |               |      |       |               |      |       |

## 4.3 Rasch Analyses of Responses to Items 4–15

Next, for the purpose of examining how perceived ability of each student and perceived difficulties of Items 4–15, which concern specific skills, changed relative to each other over time from TT01 to TT12, the Rating Scale Rasch model (Wright & Masters, 1982) was applied to the data sets at TT01 and TT12, separately, using Winsteps software (Linacre, 2005). Rasch analysis converts ordinal Likert ratings to interval logit measures, enabling one to compare person abilities and item difficulties on a single dimension (Rasch, 1960; Shizuka, 2007). In each Winsteps run, the setting was such that the mean item difficulty was fixed at 0.00 logits, to make across-administration ability comparison meaningful. The responses by 57 students who had no missing or extreme values (i.e., rating all items by 0s or 4s) in either administration were analyzed.

The mean ability at TT01 turned out to be -1.64, the standard deviation was 1.86, and person separation reliability, the Rasch equivalent of Chronbach alpha, was 0.91. At TT02, the mean was 2.77, the standard deviation was 3.08, and reliability was 0.94. A paired  $t$  test confirmed that the difference was significant ( $t = 11.38$ ,  $p = .000$ ), and the effect size was quite large ( $d = 1.785$ ), which indicates that participants' perceived abilities as measured by responses to Items 4–15 significantly and substantially improved.

**Table 4 Rasch Item Difficulties with Fit Statistics at TT01 and TT12**

| [TT01]      | Measure | SE   | Infit | Outfit | [TT12]      | Measure | SE   | Infit | Outfit |
|-------------|---------|------|-------|--------|-------------|---------|------|-------|--------|
| 15 RTM_w/oA | 2.03    | 0.25 | 1.52  | 1.43   | 15 RTM_w/oA | 2.80    | 0.26 | 1.60  | 1.65   |
| 05 R_w/oA   | 1.54    | 0.24 | 0.95  | 0.89   | 05 R_w/oA   | 2.32    | 0.26 | 0.62  | 0.55   |
| 09 DH_w/oA  | 1.13    | 0.24 | 0.75  | 0.73   | 09 DH_w/oA  | 1.69    | 0.27 | 0.58  | 0.63   |
| 07 TH_w/oA  | 1.07    | 0.24 | 0.83  | 0.84   | 13 V_w/oA   | 1.69    | 0.27 | 0.77  | 0.72   |
| 13 V_w/oA   | 0.63    | 0.23 | 0.92  | 0.89   | 07 TH_w/oA  | 1.62    | 0.27 | 0.48  | 0.44   |
| 11 F_w/oA   | 0.46    | 0.23 | 1.01  | 0.99   | 11 F_w/oA   | 1.48    | 0.27 | 0.67  | 0.65   |
| 14 RTM_wA   | 0.36    | 0.23 | 1.68  | 1.61   | 14 RTM_wA   | 0.75    | 0.27 | 1.74  | 1.95   |
| 04 R_wA     | -1.06   | 0.23 | 0.72  | 0.73   | 04 R_wA     | -2.13   | 0.33 | 0.95  | 0.96   |
| 08 DH_wA    | -1.22   | 0.23 | 0.96  | 0.95   | 12 V_wA     | -2.24   | 0.33 | 0.94  | 0.78   |
| 06 TH_wA    | -1.32   | 0.23 | 0.72  | 0.73   | 08 DH_wA    | -2.35   | 0.33 | 1.04  | 1.14   |
| 12 V_wA     | -1.78   | 0.23 | 0.88  | 0.89   | 06 TH_wA    | -2.81   | 0.34 | 0.82  | 1.26   |
| 10 F_wA     | -1.83   | 0.23 | 1.16  | 1.17   | 10 F_wA     | -2.81   | 0.34 | 1.57  | 1.30   |

Table 4 shows item difficulties (“Measure”), standard errors (“SE”) and infit and outfit mean squares (“Infit”, “Outfit”) at TT01 and TT12. Items have been sorted in the descending order from the most to least difficult, in each administration separately. It can be noted, first, that at both times all the without-attention (...\_w/oA) items were more difficult than the with-attention (...\_wA) items. Second, if the mean difficulty of each corresponding with- and without-attention pair (e.g., RTM\_wA and RTM\_w/oA) is calculated, the most difficult is stress-timed rhythm (RTM), the second most difficult is /r/ (R), and the least difficult is /f/ (F), at both TT01 and TT12. Although there were some ranking changes in the other three segmental sounds, it would be safe to say, taking standard errors into consideration, that relative difficulties of the items were practically very similar, if not identical, across administrations (Pearson’s  $r = 0.977$ ,  $p < .000$ ; Spearman’s  $\rho = .961$ ,  $p < .000$ ). This would serve as a piece of evidence for the reliability of learners’ self-claimed perceived difficulties. Another important observation is that infit and outfit mean squares of Items 15 and 14, rhythm with and without attention, are both higher than 1.5, a borderline value between good and not-so-good fits (Bond & Fox, 2001). This implies that rhythm-related items may constitute a somewhat different dimension from the other segmental items.

#### 4.5 Overall Satisfaction and Sense of Improvement

As mentioned above, an additional survey was conducted in the final session to further explore participants’ perception. The level of their satisfaction with the course turned out to be quite high. As many as 22 out of 66 students answered that they were 100% satisfied; the mean was 86.53% and the standard deviation was 14.95. With regard to self-reported degrees of improvement in pronunciation due to the course, 28 students chose “5” (“improved very much”); 27 chose “4” (“improved substantially”); 9 chose “3” (“improved a little”); 1 chose “2” (“did not improve very much”); and none chose “1” (“improved little”) or “0” (“did not improve at all”).



#### 4.6 Verbal Comments

Each participant's verbal comment in Japanese was scrutinized for occurrences of discernable propositions. Through cyclical examination, all the comments by 64 respondents (two were absent) were finally broken down into 351 propositions, which were classified into 31 categories shown in Table 7. For example, the comment (originally in Japanese):

*Practicing pronunciation was fun. I first thought a one-year training of pronunciation would be too long, but looking back now, it was all over too soon. I now believe pronunciation is important, and I am going to be careful about it when I speak English. That's because if I know how words are pronounced, I will be able to understand a foreign person speaking in English. This course was really fun. I liked the songs.*

was interpreted as containing proposition categories #05, #20, #30, #28, and #06 in Table 7. Even a quick inspection of the list is enough to reveal its overwhelmingly positive tone. Except for #03, #10, #19, and #22, the comments are expressions of the participants' pride, joy, and satisfaction. Specifically, a large number of students (40 out of 66) wrote that their pronunciation had improved (#01). A substantial number wrote that they had become more careful about pronunciation (#02), that the tasks had been difficult (#03) but beneficial (#04) and fun (#05). The use of songs and video clips proved quite popular (#06), as expected. Other noteworthy comments included: that they felt ecstatic when they finally got passes (#16); that they felt proud of what they had come to be able to perform (#09); and that they felt acquiring accurate pronunciation helped with listening to native speakers (#28).

**Table 7 Propositional Units Found in Students' Verbal Comments**

|     | Propositions   | Count |
|-----|--|-------|
| #01 | My pronunciation has improved; my rhythm is more English-like now.   | 40    |
| #02 | I have become more careful about pronunciation when reading aloud, or speaking, English.                                 | 25    |
| #03 | At first, I could not produce the right sounds; I had a hard time in the early sessions.                                 | 23    |
| #04 | The course has been beneficial; worthwhile.  | 23    |
| #05 | The course has been fun; time has passed quickly.  | 22    |
| #06 | Songs and/or video clips as training materials have been enjoyable.  | 22    |
| #07 | My understanding of English pronunciation and rhythm has deepened.   | 19    |
| #08 | I have come to like English (better than before).  | 13    |
| #09 | I felt great about what I had come to be able to do.   | 13    |
| #10 | Continually being tested was stressful; the class style was cumbersome.  | 12    |
| #11 | Continually being tested has benefited me; the class style was a lot of fun.   | 11    |
| #12 | Before the course, my pronunciation was terrible; I was not paying attention to pronunciation.                           | 11    |
| #13 | I want to make [have already made] use of the skills I have gained through the course, in English speaking environments. | 11    |
| #14 | I want to keep on doing my best in learning English.   | 11    |
| #15 | Your way of teaching made it easy to understand/acquire the target skills.   | 9     |
| #16 | I felt ecstatic when I finally got a "pass" after numerous failed attempts.  | 9     |
| #17 | I have become better at reading aloud English sentences.   | 8     |
| #18 | I have come to like reading aloud (or speaking) English (better).  | 8     |
| #19 | The tests were hard to pass.   | 7     |
| #20 | I have come to believe pronunciation is important.   | 6     |
| #21 | I want to become even better at pronunciation; I want to keep the present level of pronunciation.                        | 6     |
| #22 | I wanted to see the whole movie, not just a couple of scenes.  | 6     |
| #23 | I began to enjoy this class (after some time).   | 5     |
| #24 | I have learned how difficult it is to pronounce English words right.   | 5     |
| #25 | I enjoyed co-operative learning tasks.   | 5     |
| #26 | I have come to be able to tell the pronunciation of an unknown word from its spelling.                                   | 5     |
| #27 | I want to learn more in this course; there should be more of this type of class.   | 4     |
| #28 | My improved pronunciation skills have improved, or should improve, my listening abilities.                               | 4     |
| #29 | In my free time, I want to sing songs in English to brush up my pronunciation.   | 3     |
| #30 | I want to keep on being careful about pronunciation even after the course is over.                                       | 3     |
| #31 | Your pronunciation is just like a native speaker's.  | 2     |

#### 4.7 Correlation Analysis

In order to explore the inter-relationships between these proposition units quantitatively, an occurrence and non-occurrence of a unit was coded as 1 and 0, respectively, which produced 31 dichotomous variables. Correlations among these variables and the perceived improvement rating and the overall satisfaction, mentioned above, were computed. The correlation table is too large to be produced here, but a total of 34 significant correlations, 11 at  $p < .01$  level and 23 at  $p < .05$  level, were identified. The combinations that produced the  $p < .01$  level correlations are shown in Table 8.

**Table 8 Variable Combinations with Correlations Significant at  $p < .01$  Level**

| Variable Pair |      | $r$    | $p$ -value | Variable Pair |      | $r$    | $p$ -value |
|---------------|------|--------|------------|---------------|------|--------|------------|
| # 29          | # 21 | 0.437  | 0.000      | # 11          | # 05 | 0.374  | 0.002      |
| # 15          | # 07 | 0.430  | 0.000      | # 25          | # 24 | 0.351  | 0.004      |
| # 23          | # 08 | 0.434  | 0.000      | # 25          | # 11 | 0.333  | 0.006      |
| # 27          | # 23 | 0.407  | 0.001      | # 17          | # 11 | 0.332  | 0.006      |
| Stsfctn       | # 22 | -0.409 | 0.001      | Imprvd        | # 16 | -0.318 | 0.010      |
| # 23          | # 03 | 0.392  | 0.001      |               |      |        |            |

Some of the observable tendencies were: students who want to practice singing English songs themselves (#29) tended to be students who want to become even better (#21); those who felt the author's instructions were easy to follow (#15) tended to feel their knowledge deepened (#07); those who wrote they had come to enjoy class (#23) also wrote they had come to like English better (#08); and, understandably, those who came to enjoy class (#23) tend to write they want to learn more in this course (#27). Particularly noteworthy is that those who enjoyed being tested (#11) tended to feel the class was fun (#05), tended to enjoy working with their friends (#25), and feel they have improved in their oral reading skills (#17).

#### 4.8 Regression

Finally, to explore the degrees of contribution of different factors to participants' satisfaction with the course, a stepwise regression was conducted with the overall satisfaction percentage as the dependent variable and the 31 propositional categories plus perceived improvement ratings as independent variables. The threshold probabilities to enter/remove a variable to/from the model were conservatively set at 0.05. The obtained model, with an adjusted  $R$ -squared of 0.364 and an  $F$ -value of 10.19 ( $p = .000$ ), had four significant predictors in it (see Table 9). The model specifies that, other factors being equal, the overall satisfaction is higher when (a) the sense of improvement is stronger, (b) training in the form of testing is enjoyed, and (c) the importance of pronunciation is appreciated. (The significance of proposition #22 as a negative contributor simply means that students who had been misinformed that during the course whole movies would be watched just for fun felt betrayed, and naturally had lower levels of satisfaction.)

**Table 9 Parameter Estimates for Predicting Overall Satisfaction with the Course**

| Term              | Estimate | Std Error | $t$ Ratio | Prob>  $t$ |
|-------------------|----------|-----------|-----------|------------|
| Intercept         | 56.205   | 8.626     | 6.52      | <.0001     |
| Prcvd Imprvmnt    | 7.027    | 1.987     | 3.54      | 0.0008     |
| #11 (testing)     | 10.318   | 4.120     | 2.50      | 0.015      |
| #20 (important)   | 12.896   | 5.124     | 2.52      | 0.0145     |
| #22 (whole movie) | -27.637  | 5.376     | -5.14     | <.0001     |

## 5. Discussion and Conclusion

First, the sign tests confirmed that for almost every item, the median rating was higher at the end of the course than at the beginning, with a large effect size. It means, regarding Items

1 and 2, that the participants became more keenly aware of the importance of pronunciation and more careful about their own pronunciation when speaking English outside the course. When one's pronunciation accuracy is strongly correlated with one's concern about pronunciation (Shizuka, 1993), enhancing participants' motivation toward the skill could be regarded, by and of itself, as an important benefit of the course.

It also means their perceived abilities of the target features became much higher. Since that did not happen to the control group, the observed change could safely be attributed to what the participants experienced during the course, i.e., to the cyclical, one-on-one, performance testing. That the control group virtually did not change is consistent with the finding by Shizuka (1993) that in the Japanese EFL setting pronunciation accuracy generally does not improve over time, even when students are continually exposed to target-like pronunciation by the teacher and/or audio-materials recorded by native speakers.

The Rasch analyses corroborated the findings by the sign tests, confirming that students' perceived pronunciation abilities were markedly improved by the one-year course. In addition, relative perceived difficulties of the items were surprisingly stable across time. First, each of the with-attention items was consistently perceived to be more difficult than its without-attention counterpart. Second, whether with or without attention, the most difficult was rhythm, followed by /r/, and the least difficult was /f/. These relative difficulties in fact match the author's in-class observation of participants' performance in the PCM, which indicates that students' responses to a survey can be relied upon not only as a measure of their perception but also as an index of their actual ability. (However, this does *not* mean that such surveys can replace performance tests. The responses by the present study's participants may have been largely valid because they were told in advance that their responses would not affect their grades in any way. If they had thought their self-claimed abilities would somehow influence their grades, response validity would have been lost.) Another interesting observation was that acquiring stress-timed rhythm not only is perceived to be more difficult than segmental sounds but probably is a feature that belongs to a different dimension. That segmental and suprasegmental features follow somewhat different paths when acquired would be a plausible possibility.

Before concluding, one limitation of the study needs to be mentioned, which derives from the fact that the survey was not carried out anonymously. In a situation where respondents answered with their identities revealed to the author, there was a possibility that some consciously or subconsciously responded in a way that they believed would please their teacher. Anonymous surveys may have created somewhat different results.

However, even if the true effect sizes were somewhat smaller, it is difficult to imagine that the overall picture would be very different. Combining all the Likert scale survey results and the verbal comments, it seems fairly certain that the participants have evaluated the course quite highly mainly because they have experienced a clear sense of improvement, which they believe was brought about by the continual, one-on-one, performance testing. It is the author's hope that auditory and acoustic analyses of participant's recordings, which are currently being prepared for publication, will be able to demonstrate that improvement in participants' perceived pronunciation was a reflection of reality. Meanwhile the readers are encouraged to try out one-on-one pronunciation testing in their own classrooms to see if the

effects observed in the present case study will be replicated.

## References

- Arimoto, J. (2005). Hatuon shido ni okeru kyoshi no yakuwari [The teacher's role in pronunciation teaching]. *The English Teachers' Magazine*, 54, 27–29.
- Asami, M. & Tanaka, A. (2005). Juu dankai hatuon shido wo kensho suru [Verifying the effects of the 10-step pronunciation teaching]. *The English Teachers' Magazine*, 54, 31–33.
- Bond, T.G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. NY: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). New Jersey: Lawrence Erlbaum.
- Crystal, D. (2003). *English as a global language*. (2nd ed.). Cambridge: Cambridge University Press.
- Dauer, R. M. (2005). The lingua franca core: A new model for pronunciation instruction? *TESOL Quarterly*, 39, 543–550.
- Dickerson, L. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly*, 9, 401–407.
- Esling, J. H. & Wong, R. F. (1983). Voice quality setting and the teaching of pronunciation. *TESOL Quarterly*, 17, 89–95.
- Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30, 117–134.
- Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 15, 443–455.
- Flege, J. E., Takagi, N., & Man, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, 38, 25–55.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk and H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kosuge, K. (2005). Kihon teki na shidoho ni atarashii kuhu wo kuwaete [Add new ideas to traditional teaching techniques]. *The English Teachers' Magazine*, 54, 11–13.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.
- Labov, W. (1969). The study of language in its social context. *Studium Generale*, 23, 30–87.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39, 369–377.
- Linacre, J. M. (1997). FACETS (Version 3.0) [Computer software]. <http://www.winsteps.com/>
- Linacre, J. M. (2005). WINSTEPS (Version 3.55) [Computer software]. <http://www.winsteps.com/>

com/

- Makino, T. (2005). Naze seito wa hatuon wo machigau ka [Why do students produce wrong sounds?]. *The English Teachers' Magazine*, 54, 15–17.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481–520.
- Nakanishi, N. (2004). Japanlish hatuon ga imi rikai to taijin miryoku ni oyobosu eikyou [The effects of Japalish pronunciation on comprehension and interpersonal attraction]. Proceedings of the 7th KELES Graduation and Masters Theses Presentation Seminar, 51–56.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. (3rd ed.). New York: McGraw-Hill.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago).
- Riney, T. J., Takagi, N. & Inutsuka, K. (2005). Phonetic paramaters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly*, 39, 441–466.
- Sato, C. J. (1985). Task variation in interlanguage phonology. In S.M. Gass and C.G. Madden (Eds.), *Input in second language acquisition* (pp. 181–196). Rowley, NA: Newbury House.
- Schmidt, R. (1977). Sociolinguistic variation and language transfer in phonology. *Working Papers on Bilingualism*, 12, 79–95.
- Shizuka, T. (1993). Task variation and accuracy predictor in interlanguage phonology production. *Kanto-koshin-etsu Association of Teachers of English Bulletin*, 7, 63–80.
- Shizuka, T. (1995). Kojin kaaddo hoshiki tango risuto hatuon shidou no kouka ni kansuru jissho teki kenkyuu [Effects of Personal Card Approach to pronunciation teaching]. *Kanto-koshin-etsu Association of Teachers of English Bulletin*, 9, 11–19.
- Shizuka, T. (2007). *Kiso kara fukaku rikai suru Rasshu moderingu [Rasch modeling for objective measurement]*. Osaka: Kansai University Press.
- Shizuka, T. (2008). Backwash of pe rformance testing: Auditory and acoustic analyses of an utterance recorded before and after an EFL pronuunciation course. *JLTA Journal*, 11.
- Smith, L. E. (1983). *Readings in English as an international language*. Pergamon Press.
- Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, 233–253.
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Yamada, T., Adachi, T., & ATR Institute (1999). *Eigo supiikingu kagaku teki jotatu ho [A scientific method of improving your English speaking skills]*. Tokyo, Kodansha.