3L: The Southeast Asian Journal of English Language Studies – Vol 24(1): 112 – 127
http://doi.org/10.17576/3L-2018-2401-09

Tasks in Explicit L2 Pronunciation Instruction: FonF vs. FonFS in Improving Phonemic Accuracy and Comprehensibility

FARHAD TABANDEH
Faculty of Foreign Languages,
University of Isfahan, Iran

AHMAD MOINZADEH
Faculty of Foreign Languages,
University of Isfahan, Iran
moin@fgn.ui.ac.ir

HOSSEIN BARATI
Faculty of Foreign Languages,
University of Isfahan, Iran

ABSTRACT

The effects of focusing second/foreign language (L2) learners' attentions on phonological forms while communicating in meaningful discourse has recently attracted attention in L2 pronunciation research. One such treatment is focus-on-form (FonF) instruction wherein L2 learners practice and notice pronunciation features in communicative tasks rather than in decontextualized exercises and drills (i.e., focus-on-forms [FonFS]). Given this, the current study investigated the differential effects of FonF and FonFS instructions on improving Iranian English as a foreign language (EFL) learners' pronunciation of the most problematic English consonants. After identifying the problematic English consonants (i.e., /θ/, /ð/, /w/, /ŋ/) via remedial and expert judgment approaches, 45 pre-intermediate learners embarked on an 8-hour course. The experimental group received FonF, the comparison group received FonFS, and the control group had a free conversation class minus any feedback on the target consonants. Learners' pronunciations were measured in terms of phonemic accuracy and comprehensibility in controlled and spontaneous tasks. The results of immediate and delayed post-test for phonemic accuracy revealed that whereas both FonF and FonFS were equally effective in controlled tasks, only FonF instruction proved effective up to the delayed post-test in spontaneous tasks; no such improvements, however, were observed for the control group. Results also showed that improvements in phonemic accuracy led to overall comprehensibility enhancements in EFL learners' speech. The article concludes with some pedagogical implications of the findings.

Keywords: Comprehensibility; FonF; FonFS; L2 pronunciation; phonemic accuracy

INTRODUCTION

One of the side effects of adopting communicative approaches in second/foreign language (L2) instruction is the limited attention to pronunciation based on the assumption that the focus of L2 pedagogy needs to be on the meaning and function rather than on the form (Derwing & Munro 2005). In the past decade, however, pronunciation instruction has found a more prominent place in L2 pedagogy (Thomson & Derwing 2015). Accordingly, a revitalized attention has been paid to the effectiveness of various instructional treatments in improving L2 learners' pronunciations at both segmental and supra-segmental levels (Thomson & Derwing 2015). Nevertheless, despite the over-growing interest in embracing communicative approaches toward teaching and researching different L2 language skills, it seems that pronunciation is still mostly researched and taught in de-contextualized listen-and-repeat fashion (Celce-Murcia et al. 2010).

Most of the recent research on L2 pronunciation instruction has focused on explicit instruction of problematic L2 pronunciation features (e.g., Dlaska & Krekeler 2013, Kissling 2013, Gooch, Saito & Lyster 2016, Saito 2011a, Wipple et al. 2015). According to DeKeyser

(2003), an instructional treatment is explicit if rule explanation forms part of the instruction. One of the methodological approaches that heavily relies on explicit instruction is the focus-on-forms (FonFS) in which L2 learners produce and practice learned features in a series of decontextualized and controlled exercises and drills (Ellis 2016, Nassaji 2016). According to Ellis (2016), this type of structuralist-behaviourist instruction is opposed to the focus-on-form (FonF) instruction in which L2 learners practice and notice target features in communicative meaning-oriented activities (e.g., tasks).

Anderson (2017) argues that whereas FonFS is an instance of traditional presentation-practice-production (PPP) model (i.e., explicit instruction followed by controlled exercises and drills), FonF is an example of communicative PPP model (i.e., practice and production in communicative tasks). Based on the psycholinguistic underpinnings of skill learning theory, the FonF treatment is effective in helping L2 learners to automatize the use of target features in communicative contexts via extensive meaning-oriented practice (DeKeyser 2015).

Saito (2012) states that research in L2 pronunciation instruction has mostly focused on FonFS tradition. Literature in this regard shows that this type of instruction is effective in improving L2 learners' pronunciation accuracy and comprehensibility (i.e., ease of understanding) mostly in controlled read-aloud contexts (e.g., Dlaska & Krekeler 2013, Kissling 2013, Roohani 2013, Saito 2011a, Sturm 2013). Nevertheless, limited research has observed that FonFS cannot yield fruitful results in spontaneous communicative contexts (e.g., Saito 2011a). It appears that confining L2 pronunciation practices to form-oriented exercises and drills in lieu of communicative activities have caused difficulties for L2 learners to transfer what they have learned into communicative contexts (Celce-Murcia et al. 2010, Growther et al. 2015). Given such an overpaid attention to FonFS instruction in L2 pronunciation research at the expense of investigating the effects of FonF instruction, the following study focused on the differential effects of the two treatments on teaching problematic English consonants to Persian-speaking Iranian EFL learners.

REVIEW OF LITERATURE

A look at the pedagogical history of L2 pronunciation instruction reveals that before the dawn of the communicative language teaching (CLT) approach, there were two mainstream methodologies regarding pronunciation: intuitive-imitative and linguistic-analytic (Celce-Murcia et al. 2010). Whereas in the former learners relied on their own intuitions to imitate L2 sounds and rhythm from L2 native speakers without any instruction, in the latter learners were guided by instructors to learn and practice target features by the help of IPA charts, listen-and-repeat exercises, and contrastive analysis (Roohani 2013). It is the second type of instruction that is a direct result of technological laboratory-based advancements and the popularity of behaviouristic psychological approaches in the mid-1950s. In fact, unlike other language skills, the linguistic-analytic approach for pronunciation survived the downfall of behaviouristic audiolingual method and has continued to the present (Celce-Murcia et al. 2010).

Drawing on the tenets of structuralist-behaviouristic approaches, linguistic-analytic instruction with its emphasis on the synthetic presentation of pronunciation features in controlled contexts is a true example of focus-on-forms (FonFS) instruction (Ellis, 2016; Saito, 2012). In this type of treatment, target pronunciation features are explicitly presented to the learners and learners then engage in a series of decontextualized exercises and drills to practice the L2 features. Recent literature shows that FonFS has been the most researched and pedagogically practiced type of instruction for L2 pronunciation (Lee, Jang & Plonsky 2015, Saito 2012).

As an example, Saito (2011a) investigated the effects of explicit FonFS instruction of English problematic consonants and vowels on Japanese ESL learners' comprehensibility as well as accentedness (i.e., native-like pronunciation). Comprehensibility has been defined as how easy L2 learners' pronunciation is for interlocutors to understand (Derwing & Munro 2005, Saito, Trofimovich & Isaacs 2015). In fact, this is a holistic measurement to gauge the effects of pronunciation instruction on learners' intelligible and comprehensible speech as decided by listeners (Levis, 2005). Based on this rubric, Saito (2011a) found that this type of explicit instruction significantly improved learners' comprehensible speech in controlled read-aloud tasks in the post-test; it was not as effective in spontaneous picture description tasks, though.

The explicit instruction in FonFS tradition has also been compared with implicit instruction in which learners are required to decipher the rules themselves through mere exposure. In her study, Kissling (2013) compared these two types of instruction in helping English-native learners of Spanish to learn the problematic Spanish consonants and vowels. She found that both treatments improved learners' phonemic accuracy in controlled sentence reading tasks in the post-test.

In accordance with the above-mentioned studies, Sturm (2013) observed that FonFS instruction of L2 French segmental and suprasegmental pronunciation features could enhance the accuracy of L2 French learners in controlled reading tasks in the post-test. This finding was also later corroborated by Gooch et al.'s (2016) study on the effects of conventional explicit instruction on Japanese' learners' accuracy of English consonant /ɹ/. The mediating effects of feedback type (recast vs. prompts) and the elicitation context (controlled vs. spontaneous) were also examined. Their findings showed that those learners who received FonFS instruction followed by recasts improved only in the controlled read-aloud task, while the other group that had prompts as feedback significantly improved in both controlled and spontaneous tasks in the post-test.

Contrary to the traditional yet popular FonFS, more communicative approaches have been brought into the limelight (Celce-Murcia et al. 2010). One of the most popular recent instructions has been the use of tasks as communicative activities to follow explicit instruction. Known as focus-on-form (FonF), in this type of instruction tasks are employed to provide learners with opportunities to practice and notice already learned features in communicative contexts while the primary focus is on the meaning. A type of task that is appropriate for this purpose is 'focused task' through which learners are required to produce and/or perceive particular target features while performing the task (Ellis 2009, Celce-Murcia et al. 2010).

This type of meaning-oriented treatment has just recently attracted L2 pronunciation researchers' attention. To state an example, focusing on the problematic consonant /ɹ/, Saito and Lyster (2012) studied the role of FonF and corrective feedback in the phonetic accuracy of this consonant. With the help of acoustic analysis, they demonstrated that their treatment significantly affected learners' phonetic accuracy of /ɹ/ in controlled and spontaneous tasks in the post-test. In another study, Saito (2015) investigated how recast in FonF instruction could promote L2 pronunciation development of word-initial consonant /ɹ/ by Japanese ESL learners. His findings revealed that FonF accompanied by recast is very effective in helping learners adjust the formant frequencies of this approximant consonant. Finally, McKinnon (2017) observed that FonF instruction of L2 Spanish declarative and imperative patterns was effective in making learners observe these prosodic features in spontaneous discourse completion tasks.

In total, contemporary research shows that explicit instruction of pronunciation features followed by controlled exercises (i.e., FonFS) is mostly effective in improving learners' segmental or supra-segmental accuracy (e.g., Reis & Hazan 2011, Kissling 2013,

Sturm 2013a, 2013b) and comprehensibility (e.g., Dlaska & Krekeler 2013, Saito 2011a) in controlled contexts. Moreover, few studies have shown that adding tasks to explicit instruction (i.e., FonF) resulted in improved spontaneous pronunciation performance in immediate post-test (e.g., Gooch et al. 2016, McKinnon 2017, Saito & Lyster 2012). Nonetheless, it is still under-researched, compared to the traditional FonFS instruction, to what degree FonF instruction helps EFL learners improve segmental pronunciation accuracy in controlled and spontaneous tasks in the immediate and delayed post-tests, and to what extent such an improvement leads to more comprehensible EFL speech.

PURPOSE OF THE STUDY

As noted, FonFS instruction has been found to be effective in controlled elicitation contexts, rather than in meaning-oriented spontaneous ones (Celce-Murcia et al. 2010, Saito 2011a). Thus, if the central goal of L2 speech learning is to help learners achieve more accurate and comprehensible pronunciation in real-time L2 communicative contexts (Derwing & Munro 2005, Levis 2005, Saito & Lyster 2012) despite its grave difficulty (Al-Abdely & Thai 2016), providing learners with FonF instruction seems pedagogically advantageous (Celce Murcia et al. 2010). As asserted by Ellis (2016), unlike in the FonFS instruction wherein learners are presented with explicit instruction followed by practices and productions in exercises, in FonF instruction, learners are presented with explicit instruction followed by a series of meaning-oriented focused tasks to practice L2 features in real-life communicative contexts. Furthermore, it is argued that focused tasks could help L2 learners notice target features (Ellis, 2016) and hence proceduralize their explicit declarative knowledge of those features via ample communicative practice (Anderson 2017).

Accordingly, the current study hypothesizes that if Iranian EFL learners receive FonF instruction for the most problematic non-Persian English consonants, they may have a more accurate pronunciation performance in spontaneous tasks than those who receive FonFS instruction. Moreover, since some research in other ESL/EFL contexts (e.g., Dlaska & Krekeler 2013, Saito 2011a, Saito & Shintani 2016) show that segmental improvements can lead to more comprehensible L2 speech, it is also hypothesised that improvement in phonemic accuracy may result in a more comprehensible EFL speech. Hence, the following alternative hypotheses were used to guide the design of the study:

1. To what extent do FonF and FonFS instructions of English consonants affect Iranian EFL learners' phonemic accuracy in controlled and spontaneous tasks?
2. To what extent do improvements in phonemic accuracy of Iranian EFL learners' pronunciation affect the comprehensibility of their speech in controlled tasks?

METHOD

PARTICIPANTS

The participants of the study were 45 Iranian EFL English-majors studying in a non-state university in Tehran, Iran. All learners were in their third semester of university study with no previous record of taking a course in English phonology. They belonged to the low-intermediate proficiency level (based on their scores in the Oxford Placement Test (Allen, 2004) administered by the university), they were all L1 Persian speakers, and they did not have any experience of living in English L1/L2 countries or being instructed by an English

native teacher. The learners were randomly assigned to three groups: the experimental group (n=15) received FonF instruction, the comparison group (n=15) received FonFS instruction, and the control group (n=15) had a free conversation class with feedback on various pronunciation features but not the target consonants of the study (Table 1).

With reference to phonemic accuracy measurement, two L1 English EFL teachers (one male and one female, with the mean age of 34.1), who were also experts in English phonology, were recruited to phonemically transcribe the extracted target words produced by EFL learners. As for the comprehensibility measurement in the controlled passage read-aloud task, three Iranian EFL teachers (two males and one female, with the mean age of 31.3) rated learners' comprehensibility levels. All the teachers were remunerated for their participation in the study.

TABLE 1. Demographical characteristics and proficiency scores of the EFL leaner's

	FonF group	FonFS group	Control group
Number	15	15	15
Gender (Female/Male)	7 / 8	8 / 7	8 / 7
Age (mean/SD)	28.1 / 4.45	27.8 / 4.15	27.3 / 4.3
OPT (mean/SD)	31.4 / 5	30.7 / 5.2	29.8 / 5.3

TARGET CONSONATS

According to Ellis (2006), there are two approaches to identify difficult, problematic target language features for L2 learners: the remedial approach and the expert judgment approach. Whereas in the former researchers can rely on previous relevant findings (e.g., contrastive analysis), in the latter, researchers can refer to the experts' judgments (e.g., experienced teachers). Saito (2011b) states that combining the two approaches would lead to results that are more comprehensive. Hence, both approaches were adopted in the study.

A contrastive look at the phonological repertoire of the standard Persian and American English (Ladefoged & Johnson 2015) reveals that there are five consonants (i.e., phonemes) in English that are absent in the standard Persian: the two dental fricatives /θ/ and /ð/, the labio-velar approximant /w/, the velar nasal /ŋ/, and the alveolar approximant /ɹ/ (Persian has the alveolar thrill /r/). Thus, the contrastive analysis predicts that Persian-speaking Iranian EFL learners, even proficient ones, have difficulties pronouncing these consonants and they replace them with other Persian-existing segments (e.g., pronouncing the world *west* /west/ as /vest/). However, this does not necessarily entail that mispronunciation of these consonants heavily affects the comprehensibility of EFL speech. Thus, all these consonants were presented to 35 Iranian EFL teachers to be rated based on pronunciation difficulty (1: easy to pronounce to 9: difficult to pronounce) and instruction importance (1: unimportant to instruct to 9: important to instruct). As the results show (Table 2), four consonants (i.e., /θ/, /ð/, /w/, /ŋ/) achieved the highest scores on average and hence, were regarded as the target consonants of the study.

TABLE 2. American English consonants rated by Iranian EFL teachers out of a 9-point scale

Consonants	Pronunciation difficulty (mean)	Instruction importance (mean)	Overall rating score (mean)
/ ð /	7.5	8.4	7.9*
/ θ /	7.3	8.3	7.8*
/ w /	6.5	7.5	7*
/ ŋ /	7.6	5.5	6.55*
/ ɹ /	3.5	3.2	3.3

** Selected target consonants of the study*

INSTRUMENT

COMPREHENSIBILITY RATING

Following previous research (e.g., Dlaska & Krekeler 2013, Saito 2011a, Saito & Shintani 2016), a rating scale from 1 (very easy to understand) to 9 (very difficult to understand) was developed and given to the EFL teachers to rate the comprehensibility of learners' pronunciations. Teachers were asked to listen carefully once to the learners' recorded speech on a random basis and assign a global rating score based on how much they thought the speech was easy or difficult to understand.

MEASUREMENT TASKS

Learners' pronunciations of the target consonants were elicited and voice-recorded for phonemic accuracy in two different types of elicitation contexts based on the two opposite ends of production continuum (Lyster 2007). According to Lyster (2007), the controlled read-aloud task has the lowest cognitive processing load and learners are only required to read already presented written stimuli; on the contrary, the timed spontaneous picture description tasks has a higher cognitive processing load because learners are to produce free language within a short period in a communicative meaning-oriented context. As for the former elicitation task, for each target consonant of the study, two sentences were loaded with six words (i.e., eight sentences in total). As for the latter spontaneous task, one picture was presented to students for each target consonant (four pictures in total). Under each picture, three prompt words were written (two target words and one distractor). The learners were to describe each picture as though describing them for a person who had never seen them based on those words.

Concerning the comprehensibility rubric, learners were also required to read aloud a short story passage that included four target consonants in 16 target words. The same types of elicitation tasks were employed in the pre-test, immediate and delayed post-tests. See Table 3 for the list of target words.

TABLE 3. Target words employed in the three measurement tasks

Consonants	Measurement tasks	
	Sentence read-aloud	Timed picture description
/θ/	think, path, through, thank, thin, thought	thin, earth
/ð/	father, that, there, mother, those, this	mother, there
/w/	window, west, wiper, where, wind, without	west, window
/ŋ/	trying, sting, tongue, fangs, king, thing	king, sing
	Passage read-aloud	
	mouth, thick, path, thin, they, them, breathe, then, west, wet, went, wheel, king, tongue, sting, fangs	

MATERIALS & INSTRUCTIONAL TREATMENTS

Two types of pronunciation teaching materials were employed for the experimental (i.e., FonF) and comparison (i.e., FonFS) groups. As for the explicit phase of instruction for the FonF group, learners were presented with the articulatory (i.e., place and manner of articulation) and perceptual characteristics of each target consonant. Online posters for the place and manner of articulation were employed so that learners became familiar with how these consonants were produced. Learners were required to listen carefully to the teacher's (the first author) explanations and audio-based English-native pronunciations and produce consonants accordingly. The FonF group later had a series of focused tasks designed in a way that learners were required to produce and perceive a load of target consonants in each session. While performing the tasks, learners were also provided with explicit correction

feedback by the teacher if they had articulatory mistakes. As an example, if the learner pronounced the word *thin* as /tIn/ while communicating with their partner(s), the teacher directly mentioned the error and asked the learner to pronounce it as /θIn/.

The pronunciation focused tasks were either adopted from Celce Murcia et al. (2010) or developed by the authors based on the inherent characteristics of tasks (i.e., having a clearly-defined work-plan, focusing primarily on meaning, requiring contextualized language use, having a clearly-defined outcome, see Ellis, 2009). Consequently, the tasks were focused in nature (i.e., task completion was dependent on pronouncing and listening to the specific target consonants of the study), had the three sections in a task presentation (i.e., pre-task, task, and post-task), and included information-gap, role-play, picture storytelling and picture description-recognition types (see appendix 1 for sample materials).

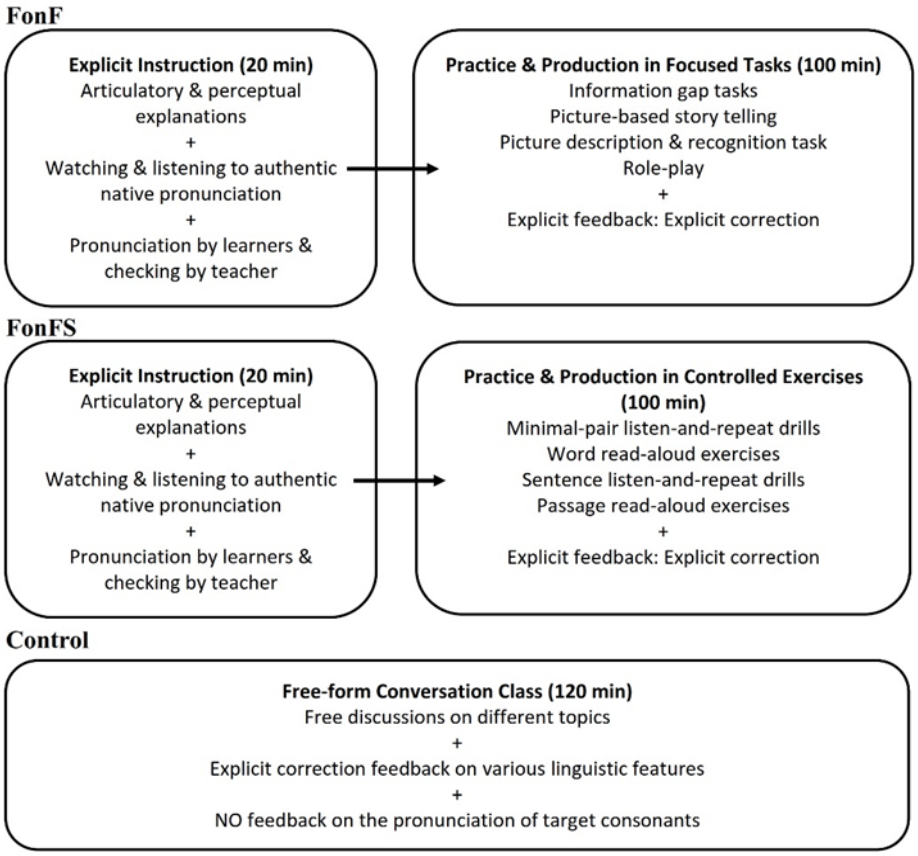

FIGURE 1. The overall instructional materials and treatments for the three groups for each session

As for the FonFS group, learners were presented with explicit instructions on the articulatory and perceptual characteristics of target consonants with the same procedure as the FonF group. Nevertheless, after explicit instructions, learners had a series of controlled decontextualized exercises and drills. The relevant chapters from Pronouncing American English textbook (Orion 2012) and Celce-Murcia et al. (2010) were employed by the teacher. The exercises and drills were minimal-pair readings, listen-and-repeat drills, and passage readings. Similar to the FonF group, the FonFS group also received explicit correction feedback. Finally, the control group only had a free discussion class on various pre-determined themes and topics at every session and received different feedback on linguistic aspects minus the target consonants of the study (see Figure 1).

DESIGN & PROCEDURE OF THE STUDY

This study had a quasi-experimental design with experimental, comparison, and control groups performing in the pre-test, immediate and delayed post-tests. After assigning EFL learners to the three groups, one week before the onset of the instruction procedure, learners were called individually to a classroom to be with the first author for their pre-test. In their pre-test, they were first required to describe four pictures based on the prompt words after only 15 seconds of looking at each picture. Afterwards, they were given one minute to look at the sentences and the passage before reading aloud these written stimuli. All the learners' productions were digitally voice recorded via a high-quality microphone and computer software.

One week after the pre-test, the instructional treatments began for FonF and FonFS groups. Learners in these two groups had two 1-hour sessions for each consonant (eight sessions in eight weeks, equalling with eight hours of instruction in total). The explicit instruction phase for each group lasted 20 minutes in each session. The rest of each session was devoted to 100 minutes production and practice in focused tasks for the FonF group and 100 minutes production and practice in controlled exercises for the FonFS group. Meanwhile, the control group had eight 1-hour free-discussion sessions.

The immediate post-tests were taken right after the last instruction session of each target consonant, and the delayed post-test was administered two weeks after the immediate post-test of each target consonant with the same procedure. The control group took the immediate and delayed post-tests in the same time intervals as the treatment groups.

DATA CODING & ANALYSIS

PHONEMIC ACCURACY

Two L1 English EFL teachers provided the phonemic transcriptions of the target consonants. Based on these transcriptions, one of the authors assigned scores to learners' productions. The human judgments of phonemic accuracy, rather than instrumental analysis, were decided sufficient because the focus of the study was on phonemic errors and phonemic contrasts (e.g., /tænk/ vs. /θæŋk/; /vest/ vs. /west/) that could affect comprehensibility and not on detailed phonetic errors reflecting allophonic variations (Ladefoged & Johnson 2015). Therefore, trained human phonemic transcriptions would easily identify the pronounced consonants. If the relevant consonant was pronounced correctly in each word, 1 point was assigned, and if it was pronounced incorrectly (i.e., substituted with another consonant), a score of 0 was assigned. This yielded two sets of scores for each participant for each target consonant. Since the results of intra-class correlation coefficient (ICC) for inter-rater reliability was substantial for all pairwise inter-rater comparisons (i.e., ICC coefficients $\geq .7$), the mean scores between the two transcribers were calculated and regarded as each learner's score. Thus, learners' scores yielded a continuous dependent variable ranging from 0 to 6 in the controlled read-aloud task and from 0 to 2 in the timed spontaneous picture description task.

After coding the scores, the results of the Shapiro-Wilk test of normality revealed that there were marked deviations from normality in the data (see Appendix 2 for the results). Therefore, non-parametric tests were employed. First, to see if there were significant differences across the three groups of the study in the pre-test, Kruskal-Wallis H test was conducted. Next, to probe any within-group improvements in the immediate and delayed post-tests, the Friedman test was conducted. If significant differences were observed within each group, Wilcoxon Signed Rank test was carried out for pairwise within-group comparisons.

COMPREHENSIBLITY

Three Iranian EFL teachers rated learners' reading of the passage based on a 9-point comprehensibility scale. Then, the ICC test of reliability was conducted to gauge the homogeneity of the given scores by three raters. The result of ICC test revealed a reliable value (i.e., .83) and the average score of each learner across three raters was calculated in each task as their final comprehensibility score.

The results of Shapiro-Wilk test of normality revealed that the comprehensibility scores displayed a normal distribution (see Appendix 3); therefore, parametric tests were employed. First, one-way analysis of variance (ANOVA) was run to gauge between-group differences in the pre-test. Later, Mixed-design ANOVA was conducted to probe overall significant within-group and between-group differences in the three testing sessions. Finally, a series of Bonferroni-adjusted post hoc analyses were run for pairwise within-group and between-group differences.

RESULTS

PHONEMIC ACCURACY

The results of Kruskal-Wallis H test revealed that there were no statistically significant differences (*P*-values ≥ .05) in the pre-test in the controlled (/θ/: $\chi 2$=.3.309, *p*=.191; /ð/: $\chi 2$=.570, *p*=.752; /w/: $\chi 2$=.703, *p*=.704; /ŋ/: $\chi 2$=.535, *p*=.765) and spontaneous (/θ/: $\chi 2$=.745, *p*=.689; /ð/: $\chi 2$=.723, *p*=.697; /w/: $\chi 2$=.011, *p*=.995; /ŋ/: $\chi 2$=1.991, *p*=.370) tasks.

To probe within-group significant improvements after the instruction, the Friedman test was employed. The results of Friedman test revealed that there were significant differences (*P*-values ≤ .05) among three testing sessions in FonF group's pronunciations in the controlled (/θ/: $\chi 2$=25.000, *p*=.000; /ð/: $\chi 2$=25.127, *p*=.000; /w/: $\chi 2$=25.857, *p*=.000; /ŋ/: $\chi 2$=27.000, *p*=.000) and picture description (/θ/: $\chi 2$=24.039, *p*=.000; /ð/: $\chi 2$=22.680, *p*=.000; /w/: $\chi 2$=25.020, *p*=.000; /ŋ/: $\chi 2$=22.333, *p*=.000) tasks. Likewise, there were significant differences in the FonFS group in controlled (/θ/: $\chi 2$=27.000, *p*=.000; /ð/: $\chi 2$=29.525, *p*=.000; /w/: $\chi 2$=27.763, *p*=.000; /ŋ/: $\chi 2$=25.481, *p*=.000) and picture description (/θ/: $\chi 2$=13.550, *p*=.001; /ð/: $\chi 2$=13.317, *p*=.001; /w/: $\chi 2$=8.340, *p*=.015; /ŋ/: $\chi 2$=18.170, *p*=.000) tasks. Nevertheless, there were no statistically significant differences for the control group neither in the controlled (/θ/: $\chi 2$=6.000, *p*=.055; /ð/: $\chi 2$=.333, *p*=.846; /w/: $\chi 2$=.154, *p*=.926; /ŋ/: $\chi 2$=1.000, *p*=.607) nor in the spontaneous (/θ/: $\chi 2$=1.600, *p*=.449; /ð/: $\chi 2$=1.500, *p*=.472; /w/: $\chi 2$=.500, *p*=.779; /ŋ/: $\chi 2$=1.500, *p*=.472) tasks.

The Wilcoxon signed-rank test was conducted for the within-group post hoc comparisons in FonF and FonFS groups (Table 4). As observed, unlike the FonF group, the FonFS group lost significant improvements over the pre-test in the delayed post-test in the timed spontaneous picture description task. In conclusion, based on these findings, the first alternative hypothesis of the study was confirmed as it was the FonF instruction that remained effective up to the delayed post-test in both controlled and spontaneous contexts. Although the FonFS yielded long-lasting instructional effects in the controlled context, it was only effective in the immediate post-test in the spontaneous context.

TABLE 4. Wilcoxon signed-ranked test's results for post hoc pairwise comparisons

		FonF group					FonFS group			
		Read-aloud		P. description			Read-aloud		P. description	
consonants	tests	Z	Sig.	Z	Sig.	tests	Z	Sig.	Z	Sig.
/θ/	pre/post	-3.415	.001	-3.364	.001	pre/post	-3.426	.001	-2.754	.006
	pre/del.	-3.432	.001	-3.241	.001	pre/del.	-3.438	.001	-.649	.516*
	post/del.	-1.228	.227	-2.200	.028	post/del.	-2.647	.008	-2.973	.003
/ð/	pre/post	-3.436	.001	-3.375	.001	pre/post	-3.421	.001	-2.753	.006
	pre/del.	-3.447	.001	-3.239	.001	pre/del.	-3.314	.001	-.141	.888*
	post/del.	-1.308	.191	-1.613	.107	post/del.	-3.451	.001	-2.541	.011
/w/	pre/post	-3.417	.001	-3.354	.001	pre/post	-3.435	.001	-2.443	.015
	pre/del.	-3.439	.001	-3.376	.001	pre/del.	-3.469	.001	-.324	.746*
	post/del.	-2.288	.022	-1.994	.046	post/del.	-3.217	.001	-2.329	.020
/ŋ/	pre/post	-3.493	.000	-3.145	.002	pre/post	-3.432	.001	-3.638	.000
	pre/del.	-3.446	.001	-3.286	.001	pre/del.	-3.395	.001	-1.134	.257*
	post/del.	-2.652	.008	-3.123	.000	post/del.	-2.887	.004	-2.652	.008

** no significant difference; P: picture*

COMPREHENSIBILITY

The results of one-way ANOVA for between-group variations in the pre-test revealed that there were no significant differences (*P*-value ≥ .05) between the three groups ($F = .717$, $p = .494$). The results of Mixed-design ANOVA for overall within-group and between-group significant differences (*P*-values ≤ .05) across the three groups and sessions showed significant differences (Table 5). The outputted partial eta squared values also proved large effect sizes for the overall within-group and between-group differences (See also Figure 2 for the relevant time plot).

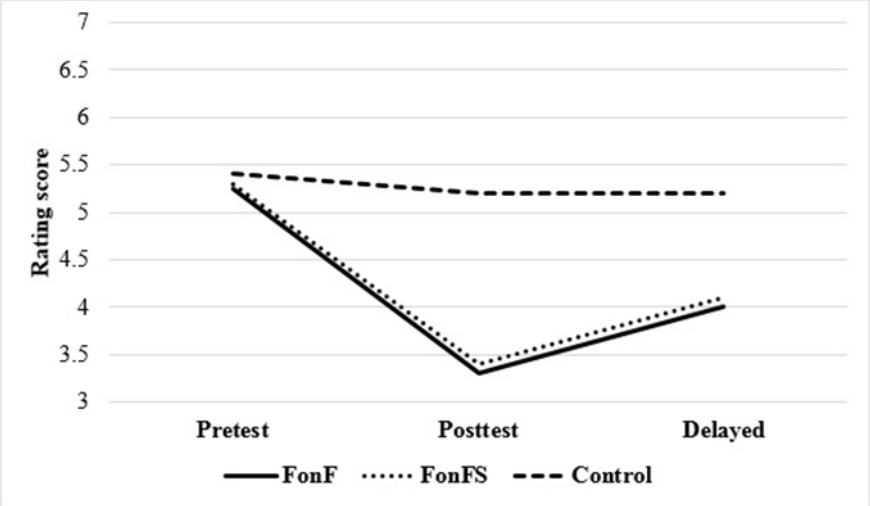

FIGURE 2. Time plot representing overall improvements in the comprehensibility scores

Nonetheless, the results of Bonferroni-adjusted post hoc within-group differences in the three testing sessions showed that, contrary to the control group with no significant improvements (*P*-values ≥ .05), both FonF and FonFS produced significantly durable instructional effects for learners' comprehensibility scores in the immediate and delayed post-tests (Table 6). Thus, in light of these findings, the second alternative hypothesis of the study is confirmed because increased phonemic accuracy of the target consonants significantly improved the comprehensibility of EFL learners' speech.

TABLE 5. The results of Mixed-design ANOVA for comprehensibility scores in passage read-aloud task

		SS	df	MS	F	Sig.	ηp2
Within-group							
	Time	42.538	2	21.269	100.392	.000	.705
	Group*Time	20.184	4	5.046	23.817	.000	.531
Between-group							
	Group	16.727	1	16.727	128.405	.037	.754

TABLE 6. The results of post hoc comparisons for within-group differences in passage read-aloud task

	FonF group	FonFS group	Control group
Testing sessions	Sig.	Sig.	Sig.
Pre-test/Post-test	.000	.000	1.000
Pre-test/Delayed post-test	.000	.000	1.000
Post-test/Delayed post-test	.000	.000	1.000

TABLE 7. The results of post hoc comparisons for between-group differences in passage read-aloud task

	Groups	Sig.
Post-test	FonF/FonFS	1.000
	FonF/Control	.000
	FonFS/Control	.000
Delayed post-test	FonF/FonFS	1.000
	FonF/Control	.000
	FonFS/Control	.000

DISCUSSION

The first objective of the study was to investigate the differential effects of FonF (i.e., explicit instruction accompanied by practices in meaning-oriented focused task) and FonFS (i.e., explicit instruction accompanied by practices in listen-and-repeat exercises and drills) on improving Iranian EFL learners' phonemic accuracy (i.e., /θ/, /ð/, /w/, /ŋ/) in controlled and spontaneous tasks. Results revealed that both types of instructions yielded significant, durable (i.e., up to the delayed post-test) improvements in learners' phonemic accuracy in the controlled sentence read-aloud context. This finding not only agrees with previous research findings on the general effectiveness of L2 pronunciation instruction (Lee et al. 2015) and explicit FonFS instruction (e.g., Kissling 2011, Rahimi & Tavakoli 2015, Sturm 2013) but also with findings in other language skills (e.g., for a discussion on explicit grammar instruction, see Spada & Tomita 2010). These results, together with the current research finding on segmental improvements, prove that explicit awareness-raising teaching of problematic language features via form-oriented practices is an effective pedagogical technique mostly in controlled elicitation contexts. As argued in skill learning theories (Anderson 2017, DeKeyser 2015), ample practice helps L2 learners proceduralize their declarative knowledge (i.e., explicitly learned features) and reach an automatized performance. Nevertheless, it seems that the nature of such a practice influences L2 learners' performance in different language production contexts (Ellis 2016).

The findings of the study demonstrated that in spontaneous communicative elicitation contexts, differential performances were observed between the two groups. Whereas in the immediate post-test, the FonF group slightly outperformed the FonFS group, the former group significantly outperformed the latter in the delayed post-test. This result highlights the role of tasks as communicative activities that could accompany explicit instruction to make L2 learners practice target features in meaning-oriented contexts. Based on the results, this type of task-supported FonF instruction (also called CLT-based PPP model by Anderson,

2017) has psycho-linguistic advantages for L2 learners. Ellis (2003) states that employing focused tasks to draw learners' attention to target forms while communicating could lead to automatization of explicit knowledge and result in implicit learning. In fact, focused tasks make learners move beyond simple controlled L2 linguistic processes and practice language use as they might face in L2 real-life L2 speech productions. Thus, FonF pronunciation instruction has the advantage over the traditional yet still popular FonFS approach to prepare EFL/ESL learners to observe segmental accuracy in communicative contexts wherein learners are under pressure to focus on the meaning and completion of the tasks (Growther et al. 2015).

On the other hand, considering the fact that immediate post-test improvements might not reflect the true amount of learning, significant improvements in the delayed post-test are better judgments of long-lasting learning (Thomson & Derwing 2015). This is an interesting finding because it shows that although particular pronunciation interventions (FonFS in this study) can lead to improvements in the immediate post-instruction performance (e.g., Saito, 2011a), they might not produce long-lasting results as measured in more distant delayed post-tests. Therefore, as argued by Thompson and Derwing (2015), treatment studies in L2 pronunciation need to employ delayed post-tests in their designs.

The second objective of the study was to investigate the relationship between improvements in phonemic accuracy of problematic constants and the overall comprehensibility of EFL learners' speech. Findings demonstrated that learners' phonemic accuracy resulted in a significant increase in comprehensibility levels of their EFL speech. In fact, the same pattern of improvement in comprehensibility was observed as in phonemic accuracy improvements in immediate and delayed post-tests. Although some researchers have asserted that segmental inaccuracy, especially with English interdentals, could not have outstanding effects on degrees of understanding by interlocutors (e.g., Jenkins 2000), the results of this study, in line with some others (e.g., Dlaska & Krekeler 2013, Saito 2011a), revealed that phonemic accuracy of interdentals can also enhance comprehensibility.

Having said that, however, the comprehensibility results should be regarded with caution since the passage was loaded with words including all four consonants of the study (i.e., /θ/, /ð/, /w/, /ŋ/). It might have been the substantial number of target words and/or the concurrent comprehensibility assessment of all four consonants in the short passage that lead to significant comprehensibility improvements. At any rate, it still could be claimed that enhancements in phonemic accuracy may cause overall comprehensibility improvements because segmental accuracy, among other variables (e.g., correct grammar and good prosody), is proven to be among the major factors affecting comprehensibility of L2 learners' speech (Saito, Trofimovich & Isaacs 2015).

In light of the above-mentioned discussion, the results of the study need to be interpreted with care because, like any other experimental research, this study had its own limitations. Primarily, the EFL learners received only two hours of instruction for each target consonant. Although this amount of treatment time seems sufficient for experimental designs, longer treatments accompanied by more distant delayed post-tests reveals more valid results of long-term improvements (see Lee et al. 2015). Moreover, this study only focused on problematic English consonants. A focus on other aspects of segmental pronunciation (e.g., English vowels) and prosody (e.g., intonation and stress) complements the results of this study on the role of task-oriented approaches in L2 pronunciation.

Regarding the measurement task for spontaneous speech production, more real-life-like elicitation tasks (such as oral discourse completion task as in McKinnon 2017) rather than timed picture description tasks can present a more naturalistic elicitation context for spontaneous L2 speech production. Finally, studying the differential effects of other types of

communicative L2 pronunciation instruction on L2 learners' segmental and supra-segmental performance may also be a promising research avenue for future studies.

CONCLUSION

The findings of this study show that traditional yet popular focus-on-forms (FonFS) pronunciation instruction is mostly effective for segmental accuracy in the controlled read-aloud context, rather than in the spontaneous communicative counterpart. In contrast, focus-on-form (FonF) instruction is not only durably effective in the controlled context but also in the spontaneous context. Additionally, in accordance with previous research findings (e.g., Dlaska & Krekeler 2013, Saito 2011a), it was observed that segmental accuracy leads to more comprehensible EFL pronunciations by the learners. Based on the delayed post-test findings, it can be argued that obtained results in the immediate post-tests may not be a valid judgment of long-lasting instructional effects because learners' performance in delayed post-tests might falter (see also Thomson & Derwing 2015).

Overall, considering the findings of the study, it can be concluded that FonF pronunciation instruction, wherein learners notice and practice pronunciation forms and meaning simultaneously, is significantly more effective than the de-contextualized FonFS instruction. Therefore, L2 pronunciation instruction needs to step away from the popular clichés of teaching and practicing the pronunciation via trivial listen-and-repeat exercises. If the central goal of ESL/EFL speech learning is to produce accurate and comprehensible L2 speech (Saito & Lyster 2012), given the integral role of pronunciation in communication breakdowns (Kang 2015), FonF instruction could provide ESL/EFL learners with opportunities to practice various pronunciation aspects in communicative meaning-oriented contexts.

REFERENCES

Al-Abdely, A. & Thai, Y. (2016). Learning English vowels by Iraqi EFL learners: Perceived difficulty versus actual performance. *3L: Language, Linguistics, Literature*, *22*(1), 1-18

Allen, D. (2004). *Oxford placement test*. Oxford: Oxford University Press.

Anderson, J. (2017). A potted history of PPP with the help of ELT Journal. *ELT Journal*, *71*(2), 218-227.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M. & Griner, J. M. (2010). *Teaching Pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, UK: Cambridge University Press.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*, (pp. 313-348). Oxford: Blackwell.

DeKeyser, R. M. (2015). Skill acquisition theory. In J. Williams & B. VanPatten (Eds.). *Theories in Second Language Acquisition: An introduction* (2nd ed.), (pp. 95-112). Mahwah, NJ: Erlbaum.

Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

Dlaska, A. & Krekeler. C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*(1), 25–37.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. New York: Oxford University Press.

Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *Applied Linguistics, 19*(1), 18-41.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246.

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research*, *20*(1), 1-24.

Gooch, R., Saito, K. & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System*, *60*(2), 117-127.

Growther, D., Isaacs, T., Trofimovich, P. & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal*, *15*(1), 80-95.

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford, England: Oxford University Press.

Kang, O. (2015). Learners' perceptions toward pronunciation instruction in three circles of world Englishes. *TESOL Journal*, *6*(1), 59-80.

Kissling, E. M. (2013). Teaching pronunciation: is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, *97*(3), 720-744.

Ladefoged, P. & Johnson, K. (2015). *A Course in Phonetics*. Hampshire: Cengage Learning Publications.

Lee, J., Jang, J. & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly*, *39*(3). 369-377.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamin.

McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declarative. *Studies in Second Language Acquisition*, *39*(2), 1-31.

Rahimi, M. & Tavakoli, M. (2015). The Effectiveness of CALL in Helping Persian L2 Learners Produce the English Vowel /ɒ/. *GEMA Online Journal of Language Studies*, *15*(3), 17-30.

Roohani, A. (2013). A comparative study of intuitive-imitative and analytic-linguistic approached to teaching pronunciation: Does age play a role? *The Asian EFL Journal, 24*(1), 87-127.

Saito, K. (2011a). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Language Awareness*, *20*(1), 45-59.

Saito, K. (2011b). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELC, 42*(3), 363-378.

Saito, K. & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 Pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*(4), 595-633.

Saito, K. & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, *50*(4), 421-446.

Saito, K., Trofimovich, P. & Isaacs, T. (2015). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization Study. *Applied Linguistics*, *36*(1), 1-25.

Spada, N. & Tomita, T. (2010). Interactions between types of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263-308.

Sturm, J. (2013). Explicit phonetics instruction in L2 French: A global analysis of improvement. *System, 41*(4), 654-662.

Thomson, R. & Derwing, T. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3)*,* 326-344.

APPENDIX 1

A SAMPLE OF TWO FOCUSED TASKS EMPLOYED FOR FONF INSTRUCTION IN THE STUDY

Task. Labio-velar /w/, listening/role-play activity.
Pre-task: Listen to someone ordering food in a restaurant. After listening for two times, try to answer the following questions. Then, ask and answer these questions with your partner.
1. What does the customer order?
2. What does he start with?
3. What does the roast chicken come with?
4. Does she order water or soda?

Task: Look at the menu on the right. Use this menu to order food. Change roles with your partner as the server and the customer.

Useful phrases for the role-play:
Customer: I'll have…… for my main course.
Customer: What does that come with?
Customer: I'd like ……., please.
Server: Anything to drink?
Server: Anything else?
Server: It comes with …….

Post-task: Now that you have practiced how to order food and how to get the order, 4 students from different groups go to board. Two-by-two practice the role-play in front of the classroom.

Appetizer:
Western soup
Squid with tomato
Warm bread
Main course:
White fish
Whale meat
Walnut stew
Beverages:
Water
White soda
Watermelon juice
Deserts:
Walnut ice cream
Wall cake
Welsh pie

Task. Interdentals (/θ/ð/), information-gap activity.

A. Pre-task: Talk to your partner about your weekly routines and schedule.

B. Task: You are presented with the weekly schedule of Brad Pit on even days (Table 1) and odd days (Table 2). You should complete Table 1's odd days by asking your partner questions, and your partner should complete Table 2's even days by asking you questions.

Example: Partner A: What does Brad Pit do on Tuesdays, at noon?
Partner B: He visits his brother.

Table 1: Brad Pit's weekly schedule on even days.

Day Period	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Morning	Thinking about future works				working on path-to-success project	Visiting her friend Catharine	
Noon			Going to theater				
Afternoon							
Evening	Watching the best movies		Visiting his father			Watching three movies	

Table 2: Brad Pit's weekly schedule on odd days.

Day Period	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Morning							
Noon		Visiting his brother		Talking to an author			
Afternoon	visiting a dentists to check teeth						Praying and thanking God
Evening		Practicing mathematics		Visiting his mother			

C. Post-task: After completion of the table based on the answers provided by your partner, report to the classroom what Brad Pit does on each day. Each student reports one day.

APPENDIX 2

THE RESULTS OF SHAPIRO-WILK TEST OF NORMALITY FOR PHONEMIC ACCURACY SCORES

		Testing sessions/Measurement tasks											
		Pre-test S.R.		Post-test S.R.		Delayed S.R.		Pre-test P.D.		Post-test P.D.		Delayed P.D	
Consonant	Groups	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.
/θ/	FonF	.663	.000	.413	.000	.930	.276	.603	.000	.870	.034	.603	.000
	FonFS	.661	.000	.561	.000	.867	.031	.716	.000	.783	.002	.561	.000
	Control	.702	.000	.581	.000	.753	.001	.581	.000	.609	.000	.649	.000
/ð/	FonF	.525	.000	.413	.000	.846	.015	.758	.001	.744	.001	.815	.006
	FonFS	.606	.000	.499	.000	.845	.015	.790	.003	.761	.001	.561	.000
	Control	.581	.000	.514	.000	.649	.000	.581	.000	.581	.000	.413	.000
/w/	FonF	.499	.000	.284	.000	.855	.057	.630	.000	.855	.057	.630	.000
	FonFS	.421	.000	.413	.000	.794	.003	.815	.006	.861	.025	.561	.000
	Control	.413	.000	.514	.000	.514	.000	.421	.000	.514	.000	.514	.000
/ŋ/	FonF	.574	.000	.542	.000	.754	.001	.556	.000	.781	.002	.633	.000
	FonFS	.601	.000	.512	.000	.791	.002	.701	.001	.615	.000	.846	.014
	Control	.546	.000	.468	.000	.833	.011	.466	.000	.790	.005	.677	.000

(S. R. = Sentence read-aloud; P. D. = Picture description; Stat. = Statistic; Sig. = P-value)

APPENDIX 3

THE RESULTS OF SHAPIRO-WILK TEST OF NORMALITY FOR PASSAGE-READING COMPREHENSIBILITY SCORES

	Testing sessions					
	Pre-test		Post-test		Delayed post-test	
Groups	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.
FonF	.938	.361	.937	.350	.902	.104
FonFS	.957	.639	.924	.223	.918	.179
Control	.791	.069	.906	.118	.917	.173

(Sig. = P-value)