



Research article

Teaching English Pronunciation at Primary School Level: A Case Study on English and Bangla Version Schools in Khulna City

Emdadul Huq, Saifullah Siddique* and Md. Ahad Hossain

English Discipline, Khulna University, Khulna-9208, Bangladesh

ABSTRACT

The role of English pronunciation in learners' communicative competence, intelligibility and fluency in speaking is very important. Although the role of oral communication in the acquisition of English is increasingly recognised, the study reveals that the teaching and learning of pronunciation is not emphasised much at the primary level in Bangladesh, especially in Bangla-version schools. This study aims to compare the pronunciation performance of students in English-version and Bangla-version schools at the primary level and to identify the factors that affect their pronunciation development in Khulna Division. A quantitative comparative case study design was used with 40 students from four schools, with two schools in the English version and two schools in the Bangla version. To ensure balanced representation, stratified and random sampling techniques were used. We administered three read-aloud tasks to the participants and recorded them using a Realme 9i smartphone. Afterwards, we analysed the recordings with SpeechAce software (an AI-powered speech assessment tool that evaluates pronunciation and speaking accuracy). The results of the pronunciation data obtained were then analysed descriptively using SPSS version 26.0 with analysis techniques of Independent Samples t-test, ANOVA, and correlation analysis. The results showed that there were highly significant differences between the two groups in terms of pronunciation accuracy, fluency, speech rate and pause frequency, favouring the students of the English-version group. The study suggests that systematic pronunciation instruction is essential for enhancing learners' oral competence. Moreover, it also recommends incorporating pronunciation-based activities and technology-assisted assessment tools into the English language curriculum in Bangladesh.

Introduction

English has become the main international language used in education, communication, technology and globalisation, and English language proficiency is becoming more and more significant in multilingual nations like Bangladesh. Pronunciation is regarded as an essential part of communicative competence among the four language skills, as intelligible pronunciation plays a crucial role in learners' communicative competence, fluency in speaking, and listening comprehension (Derwing & Munro, 2015). Successful communication in a second language depends more on intelligibility than on native-like pronunciation. Therefore, pronunciation instruction should be introduced from the early stages of second-language learning (Levis, 2018).

Children's greater phonological sensitivity and cognitive flexibility make the early stages of language

learning particularly conducive to pronunciation instruction. Providing pronunciation training at the primary school level can therefore help children develop more accurate pronunciation patterns and stronger long-term oral proficiency (Murcia et al., 2016). But English language teaching at the primary level in the context of Bangladesh is still focused on grammar-oriented and examination-based learning with less emphasis on oral communication and pronunciation practice. Consequently, many learners struggle with English pronunciation, including the production of individual sounds, stress, intonation, and fluency. These difficulties may be attributed to limited exposure to spoken English, insufficient classroom interaction, inadequate teacher training, and interference from Bangla phonology (Sultana, 2019). Although several studies have examined English pronunciation in Bangladesh, limited attention has

ARTICLE INFO

Article timeline:

Date of Submission:

16 June, 2026

Date of Acceptance:

11 October, 2026

Article available online:

15 October, 2026

Keywords:

Pronunciation teaching

Primary education

English-version schools

Bangla-version schools

SpeechAce

Bangladesh

*Corresponding author: <saifullahku20@gmail.com>

been given to primary-level learners through empirical and technology-assisted approaches. Previous research has explored pronunciation instruction, communicative language teaching, language policy, language anxiety, and classroom challenges, but technology-assisted empirical analysis at the primary level remains underexplored. Apart from this, although the differences in language exposure, method of teaching, school environment and English language use between these two versions of schools are quite significant, there is a large research gap on comparative pronunciation performance between these two school versions. This distinction is particularly relevant because English Version students generally receive greater exposure to spoken English, whereas Bangla Version students receive more instruction mediated through Bangla. These differences in language exposure may influence their pronunciation development. The present study, entitled “Teaching English Pronunciation at Primary School Level: A Case Study on Two English and Bangla Version Schools at Khulna”, is designed to explore the nature of pronunciation teaching practices and to compare the pronunciation performance of the learners of primary-level English and Bangla version schools in Khulna using SpeechAce software analysis. The study particularly aims to find out what the common pronunciation problems are, compare the pronunciation accuracy and fluency, and investigate the influence of various education systems on young learners' pronunciation skills in English. The research aims to make a contribution in the following fields: pronunciation pedagogy, teacher awareness and the enhancement of primary-level English language education in Bangladesh.

Literature Review

The importance and theories of English pronunciation

Pronunciation has received increasing attention in second-language (L2) research because the ability to produce intelligible L2 speech is closely associated with communicative competence, listening comprehension, and speaking fluency. The current approach to pronunciation teaching is one of intelligibility, rather than native-like pronunciation, with the aim of ensuring that listeners can understand the speaker (Levis, 2020). Recent pronunciation studies indicate that although learners might have solid grammatical competence, they still may not be able to communicate effectively and comprehensibly because of their pronunciation (Stoughton & Kang, 2024). In contemporary theory, therefore, the two types of features were considered — segmental features (such as vowels and consonants) and suprasegmental features (such as stress, rhythm, connected speech, and intonation). Suprasegmental features play a particularly important role in intelligibility, as misplaced stress, excessive pausing, and inappropriate intonation can affect intelligibility more substantially than minor segmental errors (Munro & Derwing, 2020). The Intelligibility Principle and Communicative Language Teaching (CLT) approach consequently advocate meaningful communication, interaction, and comprehensible speech instead of accent perfection. The CA Hypothesis is also of great significance in multilingual contexts like Bangladesh, as the analysis of L1 phonology has a significant influence on L2 pronunciation. The learners who speak Bangla often bring

their native phonological properties into English pronunciation, causing problems in the vowels, consonants and stress patterns. Hence, recent research in pronunciation has highlighted the need for systematic pronunciation instruction, class interaction, corrective feedback, and technology-based learning to enhance learners' oral communicative competence and intelligibility (Metruk, 2024).

The pronunciation teaching and challenges at the primary level

Pronunciation instruction at the elementary school level can be particularly effective because children are generally more sensitive to phonological patterns and more flexible in developing phonological representations than adults. The research on second language acquisition indicates that early pronunciation teaching can help children establish more accurate pronunciation habits, enhance listening skills, and boost their oral communication confidence (O'Brien, 2021). But there are several factors that have to be taken into account for pronunciation acquisition to be effective: amount of exposure to spoken English, speaking opportunities in the classroom, teacher competence, and opportunities for communicative practice. In many cases (English as a Foreign Language (EFL) settings), pronunciation is one of the least focused areas in L2 pronunciation instruction, since the emphasis placed on pronunciation in the educational system is often to improve the ability to read and to perform well on an examination, rather than on becoming an effective communicator. This is particularly evident in the context of Bangladesh, where English language instruction at the primary level is often focused on written examination and not on speaking and listening. Learners from Bangladesh, therefore, tend to have a problem with the production of vowels and consonants, word stress, rhythm and fluency, and intonation as they lack oral practice and exposure to authentic spoken English. Mother tongue interference with the Bangla phonology causes these pronunciation difficulties to be more acute since some of the phonemes in Bangla are absent in the English phonetic system. Learners often replace unusual English sounds with familiar Bangla sounds, making the word difficult to understand. Learners frequently replace unusual English sounds with familiar Bangla sounds, which makes the word hard to understand. Recent pronunciation studies also show that teachers are not sufficiently prepared and confident in their own pronunciation knowledge to systematically teach pronunciation, leading to less frequent and less effective classroom pronunciation activities (Cengiz, 2023). Besides, tight classrooms, lack of technical facilities and communicative interaction pose further obstacles to the pronunciation development of young learners. Hence, scholars have recommended integrating pronunciation-focused activities, communicative speaking tasks, and pronunciation technology tools into primary English language learning to enhance learners' intelligibility and fluency (Tseng et al., 2022).

Pronunciation Assessment using English Language Technology in Bangladesh

Globalisation, technological development and the increasing significance of English in the educational and employment fields have introduced a lot of changes in English language teaching in Bangladesh. Communicative Language Teaching (CLT) has officially been integrated into the national curriculum, but many schools continue to use teacher-centred, examination-oriented teaching practices that offer limited opportunities for students to interact in the target language and develop their pronunciation. There are significant variations between English-version and Bangla-version schools in Bangladesh in language exposure, classroom communication, methodology and the use of English in educational activities. English-version schools provide better opportunities for English-based learning, English-based communication in lessons, English-based multimedia and English-based communicative speaking practice, which plays a positive role in affecting learners' pronunciation and fluency. In contrast, in the Bangla version, school education is more focused on Bangla-mediated teaching, and the teacher is more inclined to use a grammar-based teaching method, which reduces opportunities for students to engage in real-time oral communication in English. Such differences in education can have a profound impact on students' pronunciation, rate and fluency. In recent years, pronunciation assessment with the aid of technology has become a more important topic as digital pronunciation tools are more objective, detailed and immediate than traditional pronunciation assessment in the classroom. Automatic Speech Recognition (ASR) and Computer-Assisted Pronunciation Training (CAPT) systems are able to systematically and consistently assess pronunciation accuracy, fluency, stress patterns, pauses and speech intelligibility. (Nickolai, 2024). Mobile-assisted and ASR-based pronunciation tools are effective in assisting pronunciation development in recent systematic reviews, as they offer individualised corrective feedback and measurable speech analysis (Stoughton & Kang, 2024). One of the noteworthy example of technology-assisted pronunciation assessment is the SpeechAce software that was used in the present study, which measures fluency, pronunciation accuracy, pacing and oral clarity of the learners' pronunciation through automated measurements. Thus, it can be concluded that software-based pronunciation analysis is a suitable, methodologically sound approach to the comparison of the pronunciation performance of English-version and Bangla version primary school students of Khulna.

Methodology

This study took a quantitative, comparative case study approach to explore the pronunciation performance of the primary level learners of English in the two types of schools, namely English-version and Bangla-version schools in Khulna. A quantitative approach was chosen because the performance of the children was quantified using the scores generated by the software; therefore, an objective statistical comparison was possible between the two educational streams. When researchers want to analyse measurable variables systematically using

statistical procedures, quantitative research is deemed appropriate (Creswell & Creswell, 2018). As the study aimed to look at the differences in the pronunciation performance between the learners of English and Bangla versions in one and the same sociolinguistic context, a comparative case study design was used. A total of 40 students were selected from 4 schools, with 10 students in each school. To make sure the same number of schools from both education streams, stratified sampling was used, initially splitting schools into two strata: the English-version schools and the Bangla-version schools (the locations of the schools are shown in *Figure 1*). Stratified sampling can be useful in enhancing representativeness and reducing sampling bias in a population that consists of distinct sub-populations (Taherdoost, 2016). Then, two schools were randomly selected from each category to keep the researcher's bias to a minimum and to be as objective as possible in selecting schools. Each school provided 10 participants, of whom 5 were from Class Four and 5 from Class Five were selected, with a view to representing the two adjoining primary level classes to ensure sample representativeness and to provide a chance to observe the pronunciation performance of learners of similar age groups. Data were gathered by using a reading aloud task where the participants were asked to read three texts with commonly problematic English vowels, consonants, stress patterns, and connected speech features, which were suitable for Bangla-speaking learners. The use of reading-aloud tasks in pronunciation research is commonplace because they generate controlled speech samples appropriate for both phonological and fluency analysis (Derwing & Munro, 2015). The participants' oral productions were recorded separately to ensure clear audio without losing the meaning of the participants' language and to facilitate access to the oral productions in the classroom, using the Realme 9i smartphone. Thanks to the digitalisation of mobile devices since the last century, audio recording with a smartphone has become more and more common in modern pronunciation studies, as smartphones are considered reliable in recording speech data for acoustic and pronunciation analysis (Metruk, 2024). Due to the fact that technology-assisted pronunciation analysis is more objective, systematic and consistent than traditional human judgment (Garcia et al., 2020; Evers & Chen, 2021), the recorded speech samples were then analysed using the Automatic Speech Recognition (ASR) based pronunciation assessment software called SpeechAce. SpeechAce provided pronunciation accuracy, fluency, speech rate, and pause frequency scores for the accuracy of the two learner groups. Lastly, the quantitative data that were obtained from SpeechAce were analysed with the aid of the Statistical Package for the Social Sciences (SPSS) version 26, which were further processed using descriptive and inferential statistical analysis. The choice of SPSS was based on the fact that it is one of the most widely used statistical software packages in social science and applied linguistics research, which would enable the efficient organisation, analysis and interpretation of quantitative data (Pallant, 2020).

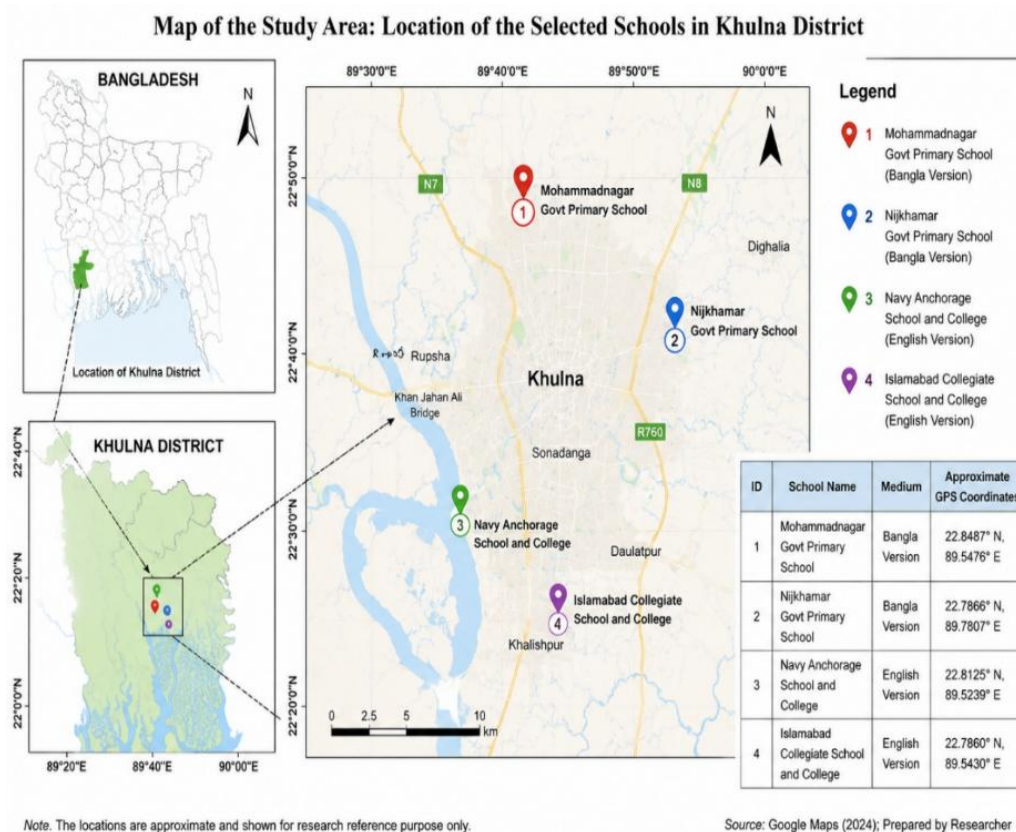


Figure 1: Locations of the Schools

Result

To analyse the collected data, several statistical tests were conducted using SPSS. Descriptive statistics were employed to summarise the respondents' characteristics and overall response patterns. Pearson correlation analysis was performed to examine the relationships among the

study variables, while independent samples t-tests were used to identify differences between groups. Additionally, one-way ANOVA was conducted to determine whether statistically significant differences existed across multiple groups. The results of these analyses are presented in the following tables.

Table 1: Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
pronunciation Rate (SpeechAce)	40	3.0	7.7	5.065	1.1718
fluency Rate (SpeechAce)	40	4.4	7.3	5.645	.7746
Avarage Accuracy	40	33.0000	85.0000	56.21666	13.04442
Total Word per min	40	32	389	171.82	102.328
Avarage bad Pause	40	.33333	25.3333	10.6500	7.07668
Valid N (listwise)	40				

The descriptive statistics give the general characteristics of the speaking performance of the 40 participants on 5 variables: pronunciation rate, fluency rate, average accuracy, words per minute and bad pauses. The mean pronunciation rate obtained was 5.07 (SD = 1.17) on the SpeechAce scale, ranging between 3.0 and 7.7, which shows that there is a moderate level of pronunciation with some variations among the learners. The mean fluency rate was slightly higher, 5.65 (SD = 0.77), ranging from 4.4 to 7.3, indicating that they were generally more fluent than accurate in pronunciation and that fluency scores were fairly uniform across learners. The overall mean accuracy score was 56.22% (SD = 13.04), ranging from a minimum of 33% to a maximum of 85%, indicating moderate accuracy of the participants' speech with considerable differences among them. As for

speaking speed, the participants' average was 171.82 words per minute (SD = 102.33) and ranged widely from 32 to 389 words per minute, showing a wide range of speaking rates. Lastly, the mean number of bad pauses was 10.65 (SD = 7.08), ranging from 0.33 to 25.33, which indicates that many learners experienced hesitation and interruptions during their speech, but the number of these pauses was somewhat different across learners. In general, the descriptive characteristics reveal moderate pronunciation, fluency and accuracy levels with significant differences in speaking speed and hesitation between individuals. This was done to see if there were any significant differences between the groups in pronunciation rate, fluency rate and average accuracy. The results revealed that the significance value was .043, which is lower than the level of significance of .05; the groups

had a statistically significant difference in their pronunciation performance. Likewise, the average accuracy score was significantly different among the groups, with a significance value of .022, which is less than .05. The difference in fluency between the groups was not

statistically significant, however, with a significance value of .052, which is close to .05. In general, the results found that there were significant differences in pronunciation and accuracy between the groups, while there was no significant difference in fluency.

Table 2: ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Pronunciation Rate (SpeechAce)	Between Groups	10.721	3	3.574	3.004	.043
	Within Groups	42.830	36	1.190		
	Total	53.551	39			
Fluency Rate (SpeechAce)	Between Groups	4.469	3	1.490	2.833	.052
	Within Groups	18.930	36	.526		
	Total	23.399	39			
Average Accuracy	Between Groups	1544.856	3	514.952	3.641	.022
	Within Groups	5091.267	36	141.424		
	Total	6636.122	39			

Table 3: Correlations

		Fluency Rate (SpeechAce)	Pronunciation Rate (SpeechAce)	Average Accuracy	Average Bad Pause	Average words per min
Fluency Rate (SpeechAce)	Pearson Correlation	1	.904**	.906**	-.687**	.947**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	40	40	40	40	40
Pronunciation Rate (SpeechAce)	Pearson Correlation	.904**	1	.990**	-.667**	.967**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	40	40	40	40	40
Average Accuracy	Pearson Correlation	.906**	.990**	1	-.693**	.976**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	40	40	40	40	40
Average bad Pause	Pearson Correlation	-.687**	-.667**	-.693**	1	-.757**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	40	40	40	40	40
Average words per min	Pearson Correlation	.947**	.967**	.976**	-.757**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	40	40	40	40	40

A correlation test was used to explore the relationships between fluency rate, pronunciation rate, average accuracy, average bad pauses, and average words per minute. The findings revealed a high positive correlation between fluency, pronunciation and accuracy in words per minute. This indicates that the students' pronunciation and accuracy also became better when they spoke more fluently, and they were able to speak a more number of words per minute. For instance, there were very strong

positive correlations between fluency and pronunciation ($r = .904$), accuracy ($r = .906$), and words per minute ($r = .947$). Likewise, there was a high correlation between pronunciation and accuracy ($r = .990$) and words per minute ($r = .967$). However, negative correlations were seen with all the other variables for average bad pauses. This suggests that the students with more bad pauses had lower fluency, less pronunciation, less accuracy and fewer words per minute. All significance values were .000, which indicates that the relationships were not random, but statistically significant.

Table 4: Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Pronunciation Rate (SpeechAce)	Equal variances assumed	.270	.610	1.515	18	.147	.7300	.4819	-.2824	1.7424
	Equal variances not assumed			1.515	18.000	.147	.7300	.4819	-.2824	1.7424
Fluency Rate (SpeechAce)	Equal variances assumed	3.133	.094	1.425	18	.171	.5500	.3860	-.2611	1.3611
	Equal variances not assumed			1.425	16.020	.173	.5500	.3860	-.2683	1.3683
Average Accuracy	Equal variances assumed	.171	.684	1.497	18	.152	7.966666666666666	5.320458998114476	3.211202906670479	19.144536240003802
	Equal variances not assumed			1.497	17.999	.152	7.966666666666666	5.320458998114475	3.211243846585566	19.144577179918890
Average bad Pause	Equal variances assumed	.001	.970	-2.278	18	.035	4.833333333333333	2.121989516037447	9.291467876736696	-.375198789929974
	Equal variances not assumed			-2.278	18.000	.035	4.833333333333333	2.121989516037447	9.291467930785700	-.375198735880971
Average words per min	Equal variances assumed	.423	.524	1.774	18	.093	24.666666666666666	13.903228153433280	4.542931791380198	53.876265124713540
	Equal variances not assumed			1.774	17.974	.093	24.666666666666666	13.903228153433279	4.545978072330943	53.879311405664280

To compare the speaking score of two groups for pronunciation rate, fluency rate, average accuracy, average bad pauses and average words per minute, the independent-samples t-test was used. Levene's Test for Equality of Variances showed that the assumption of equal variances was met for all variables, as all significance values were above .05. The results showed that there were no significant differences between the two groups in terms of pronunciation rate ($p = .147$), fluency rate ($p = .171$), average accuracy ($p = .152$), and average words per minute ($p = .093$), which means the groups did not differ significantly in these aspects of speech. A statistically significant difference was found in the average bad pause ($p = .035$), with one group having fewer unnecessary pauses than the other. The result of -4.83 in the mean difference value suggests that there is a difference between the two groups, with one group having smoother speech with fewer hesitations. Generally speaking, the results showed that, besides bad pausing, the two groups had comparable performances in all other speech performance tasks, with one group showing greater control and speech flow in the bad pausing task.

Discussion

The descriptive statistics revealed that the learners' level of fluency and pronunciation was moderate, and the learners' accuracy, speed of speech and hesitation patterns were moderate, showing that there were moderate differences among the learners. This result is in line with those of previous studies, which indicated that EFL learners may

attain a semi-functional proficiency in their spoken skills without being accurate in phonology or fully in control of speech. Derwing and Munro (2009) claim that second-language speakers can become fluent without attaining high levels of pronunciation accuracy because fluency and pronunciation are two aspects of oral proficiency. Likewise, Skehan (2009) claims that students often pay more attention to delivering a message than to accuracy, and therefore, although they have problems in pronunciation and accuracy, they do not hesitate to use the language with moderate fluency. In addition, the significant difference in speech rate and pauses found in the present study also corroborates Segalowitz's (2010) assumption that oral fluency is closely tied to cognitive automaticity that varies in acquisition among learners as a result of different exposure and practice opportunities. There are some scholars, however, who question the idea that moderate pronunciation and fluency equate with low levels of speaking competence. In a cross-cultural and multi-linguistic setting, for instance, Jenkins (2007) maintains that intelligibility trumps native-speaker norms of pronunciation. In this context, a learner can be judged a successful speaker regardless of the level of his/her pronunciation marks. Similarly, Levis (2018) argues that pronunciation assessment sometimes focuses too much on native-like pronunciation and sometimes fails to accurately reflect learners' communicative competence. Moreover, Kang's (2010) study indicates that the perceptions of listeners on the competence of speakers are not only related to pronunciation accuracy and speed of speech, but also to interactional competence and

organisation of speech. The thinking behind these points suggests that scores in these categories should not be equated with low-level oral proficiency. However, the results of the present study can be justified in the context of EFL in Bangladesh. The medium level of pronunciation and fluency, as well as the noticeable differences in speech pace and hesitation in the utterances, are a reflection of the educational facts highlighted by Hamid and Baldauf (2008), who report that English language teaching in Bangladesh may not provide much space for real-life oral communication, though learners' English learning is heavily stressed. Likewise, Ali and Walker (2014) discovered that many Bangladeshi learners have good theoretical knowledge of English but are facing problems in spontaneous spoken interactions due to less speaking practice and examination-oriented classroom teaching. Thus, the differences in the level of speech performance that were observed may be attributed to the different levels of exposure to spoken English and communicative experience of learners, and therefore are credible and reasonable in theory and context. The results of ANOVA indicate that there were statistically significant differences among the learners' groups in terms of pronunciation and average accuracy, while fluency was not significantly different across these groups. The significant difference in pronunciation performance ($p = .043$) and accuracy ($p = .022$) justifies the argument of Celce-Murcia et al. (2010) that second language learners' pronunciation competence is not equal due to differences in exposure, instruction and phonological awareness. Likewise, Gilakjani (2016) reported that there are significant variations in the accuracy of pronunciation when learners are learning English in EFL classrooms due to mother-tongue interference and limited opportunities for authentic oral interaction. However, the fact that there was no significant difference in fluency ($p = .052$) contradicts earlier research like that of Tavakoli and Skehan (2005), claiming that fluency is very context-dependent and tends to vary significantly across different learner groups. The present finding, on the other hand, indicates that learners might acquire a comparable degree of speech flow while having varied levels of pronunciation precision and accuracy. This may be attributed to several aspects of the communicative practices that are prevalent in Bangladeshi classrooms, in which students are frequently asked to keep talking even though they make mistakes in pronunciation. Furthermore, in contemporary language instruction, intelligibility and communicative competence are often more important than native-like pronunciation, which may account for the fact that the groups did not differ significantly in fluency, and that some groups' pronunciation and accuracy were quite different from others (Levis 2005). Thus, these findings suggest that learners' phonological precision and the accuracy of their speech are more sensitive to educational and linguistic differences than fluency. A correlation analysis showed strong and statistically significant relations between the various aspects of speaking performance, such as fluency, pronunciation, accuracy, words per minute and bad pauses, indicating that these are closely related. There are high positive correlations among fluency, pronunciation, and accuracy, suggesting that those who were more fluent were more accurate and more fluent in pronunciation. This result is in line with the

opinion of Lennon (2000), who claimed that fluency is closely related to linguistic competence and efficient speech processing in L2 communication. Likewise, De Jong and Perfetti (2011) demonstrated that there is a strong relationship between faster speech rate and automaticity on the one hand, and pronunciation accuracy and oral proficiency on the other. The strong association between pronunciation and accuracy ($r = .990$) also supports Isaacs and Trofimovich (2012), who suggested that a relationship between pronouncing ability and listeners' judgements of competence and communicativeness exists. But some researchers have argued that fluency is not necessarily a measure of pronunciation. Riggensbach (1991), for instance, said that students can sometimes be successful in communication despite producing less than perfect pronunciation skills, and that this may be due to learners' lack of fluency. In this light, fluency and pronunciation are not always a match. However, the present study reveals the fact that the learners' speaking skills improved in a highly interrelated way and improvement in one skill was bound to be accompanied by improvement in other skills of the learners. The other findings agree with Kormos and Dénes (2004), who stated that excessive hesitation makes speech incoherent and diminishes the perception of oral proficiency, as evidenced by the negative correlations between bad pauses and all other variables of the speech. Based on the findings, it may be concluded that hesitation, pronunciation, fluency and speaking speed are highly interconnected in the aspect of oral communicative competence for primary-level EFL learners from Bangladesh. Based on the results of independent samples t-tests, the two learner groups were not different in terms of pronunciation, fluency, accuracy, and speaking speed, but only in bad pauses. Although there may have been some variation in the educational levels, the absence of any statistically significant differences in the pronunciation rate, fluency rate, average accuracy and words per minute shows that both groups had relatively similar levels of overall oral proficiency. This discovery is in line with the argument of MacIntyre et al. (1998), which stated that besides the instructional context, the psychological and communicative aspects of a learner, including confidence, willingness to communicate and anxiety, also played a role in determining the oral performance of the learner. Roach (2009), on the other hand, noted that second-language learners can attain similar fluency and pronunciation even under different second-language learning contexts, as oral competence is built up through repeated communicative practice. This finding, however, is in contradiction with some of the studies that indicated that there are significant differences in spoken English performance between Bangla medium and English medium learners, attributed to exposure, classroom interaction, communicative resources, etc. (Hasan and Akhand, 2009). However, the present study indicates that such differences may not necessarily strongly affect core speaking measures of primary-level learners. However, the statistically significant difference between bad pauses suggests that one group had improved speech continuity and some control of hesitation. The results are consistent with Hughes (2011), who reported disfluencies like pauses and hesitations as a significant marker of processing difficulty and lack of automaticity in second-language speech

production. Thus, in spite of the similarity of the two groups on most measurable aspects of speaking, their hesitation patterns suggest that speaking flow and real-time language processing skills are still valid measures of oral communicative competence.

Conclusion

The present study examined the pronunciation performance of primary level students from English version and Bangla version schools in Khulna with the help of pronunciation analysis using the SpeechAce tool and found that there were significant differences between the pronunciation accuracy, fluency, speech rate and pause frequency of students in English version and Bangla version schools. The findings revealed that learners from schools offering English language instruction performed better in terms of pronunciation-related aspects, as they had more opportunity for listening to English, had a more communicative approach in their classroom instruction, and had comparatively stronger environments in which they learned English than the learners of Bangla-version schools did. Conversely, the learners of Bangla-version schools showed comparatively lower performance in pronunciation as they had limited opportunity for listening to the English language, had a comparatively stronger environment for learning the Bangla language, had a comparatively less communicative approach in their classroom instruction, and had comparatively fewer pronunciation-based classroom activities and a predominantly grammar-oriented approach in education. The results obtained from statistical analyses using Independent Samples t-test, ANOVA and correlation analysis also reinforced the aforementioned findings that there is a close relationship between pronunciation ability and other aspects of communicative competence, such as fluency, speech pacing and oral confidence; thus, pronunciation is a fundamental tool in the development of overall communicative competence of young learners. It was also found that pronunciation teaching is still somewhat underdeveloped in the primary education system of Bangladesh, even though it is a crucial skill in effective oral communication and intelligibility. This study strongly recommends systematic pronunciation instruction at the primary level in the education system of Bangladesh, as it is one of the best ages to master accurate pronunciation, when the student is more phonologically

flexible and sensitive to the sounds of a language. Reading aloud, phonetic awareness activities, stress and intonation practice, listening activities, classroom interaction, and speaking tasks should be introduced from the beginning of English instruction to help students gradually master correct and accurate English pronunciation. Additionally, teachers need adequate pedagogical skills to teach pronunciation, as many are under-trained or lack confidence in teaching it in the classroom. Besides, the study suggests the introduction of technology-assisted pronunciation tools and Automatic Speech Recognition (ASR) based tools in schools, as they can objectively evaluate pronunciation, give immediate feedback, and make students aware of their pronunciation problems. Based on the above results, it may be concluded that the Bangladesh education system requires a change in the way grammar is taught and needs to put more emphasis on oral communication and pronunciation development from the primary level. This would enhance learners' pronunciation accuracy and fluency and boost learners' confidence and communicative competence in English and their future academic and professional development in English language contexts.

Acknowledgement

The authors are grateful to the participants of different schools in Khulna City and the school authorities for their cooperation. The authors report that there is no conflict of interest.

Authors' Contributions:

Md. Emdadul Huq: Supervision, Questionnaire Preparation, Review, Editing, Conclusion and Recommendations.

Saifullah Sddique: Writing final draft, Methodology, Literature Review, Theoretical Framework, Data Collection. Md. Ahad Hossain: Data Analysis and Discussions;

Funding Statement: This research received funding from a grant from Khulna University Research Cell.

Conflict of Interest

The author declare no conflict of interest.

References

- Ali, M. M., & Walker, A. L. (2014). 'Bogged down' ELT in Bangladesh: Problems and policy. *English Today*, 30(2), 33–38. <https://doi.org/10.1017/S0266078414000108>
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.
- Cengiz, B. C. (2023). Computer-assisted pronunciation teaching: An analysis of empirical research. *Participatory Educational Research*, 10(3), 1–18.
- De Jong, N. H., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533–568. <https://doi.org/10.1111/j.1467-9922.2010.00620.x>
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490. <https://doi.org/10.1017/S026144480800551X>
- Gilakjani, A. P. (2016). English pronunciation instruction: A literature review. *International Journal of Research in English Education*, 1(1), 1–6.
- Hamid, M. O., & Baldauf, R. B. (2008). Will CLT bail out the bogged-down ELT in Bangladesh? *English Today*, 24(3), 16–24. <https://doi.org/10.1017/S0266078408000254>

- Hasan, M. K., & Akhand, M. M. (2009). Challenges and suitability of TESL at the college level in Bangladesh. *Journal of NELTA*, 14(1–2), 45–54.
- Hughes, R. (2011). *Teaching and researching speaking* (2nd ed.). Pearson Education.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. <https://doi.org/10.1016/j.system.2010.01.005>
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. <https://doi.org/10.1016/j.system.2004.01.001>
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). University of Michigan Press.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.
- Levis, J. M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, 6(3), 310–328. <https://doi.org/10.1075/jslp.20050.lev>
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Metruk, R. (2024). Mobile-assisted language learning and pronunciation instruction: A systematic literature review. *Education and Information Technologies*, 29, 16255–16282. <https://doi.org/10.1007/s10639-024-12453-0>
- Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, 6(3), 283–309. <https://doi.org/10.1075/jslp.20038.mun>
- Nickolai, D. (2024). Quantifying the impact of ASR-based instruction: What does the iSpraak platform learner data show? *The EuroCALL Review*, 32(2), 45–61. <https://doi.org/10.4995/eurocall.2024.20221>
- O'Brien, M. G. (2021). Ease and difficulty in L2 pronunciation teaching: A mini-review. *Frontiers in Communication*, 5, Article 626985. <https://doi.org/10.3389/fcomm.2020.626985>
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423–441. <https://doi.org/10.1080/01638539109544795>
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Stoughton, A. M., & Kang, O. (2024). A systematic review of empirical mobile-assisted pronunciation studies through a perception–production lens. *Languages*, 9(7), Article 251. <https://doi.org/10.3390/languages9070251>
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins Publishing Company.
- Tseng, W.-T., Chen, S., Wang, S.-P., Cheng, H.-F., Yang, P.-S., & Gao, X. A. (2022). The effects of MALL on L2 pronunciation learning: A meta-analysis. *Simulation & Gaming*, 60(5), 1–24. <https://doi.org/10.1177/07356331211058662>