Linguistic Forum

Published by
MARS Publishers

ISSN (Online) 2707-5273
Volume 5, Issue 2, 2023
http://doi.org/10.53057/linfo/2023.5.2.4
Pages 22-31

OPEN ACCESS

The Role of Diacritics in Reading Urdu. Can children read without "the dots"?

Correspondence:

Maria Isabel Maldonado
<spanishprofessor1@gmail.com>
Incharge/Professor, Institute of Languages and Linguistics, University of the Punjab, Punjab, Pakistan.

Muzmmal Haider
<muzmmal.haider@ell.uol.edu.pk>
Lecturer, Department of English Language and Literature, University of Lahore, Punjab, Pakistan.

Publication details:

Received: April 10, 2023 **Accepted:** May 28, 2023 **Published:** June 30, 2023

Abstract

Diacritic marks are present in numerous languages, performing diverse roles as they influence both semantics and phonetics. While they might be disregarded in certain instances of writing, they play a pivotal role in distinguishing homographs, thereby presenting challenges for L2 learners, children learning to read, and individuals with disabilities. The present study delves into the impact of diacritics on children's reading of Urdu textbooks, employing an experimental research approach wherein two sample groups (experimental and control) underwent pre and post-tests. Furthermore, the Error Analysis Model proposed by Ellis (1994) was applied for error analysis. The findings demonstrate that reading texts with diacritics proves to be easier for children. The study aims to aid Urdu textbook writers in addressing reading difficulties faced by children due to the absence of diacritic marks.

Keywords: reading problems, reading policy, diacritics, diacritics of Urdu language

1. Introduction and Background

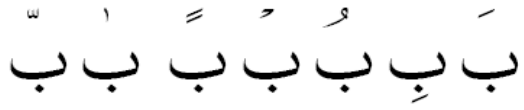
Urdu, acknowledged as the national and official language of Pakistan (Maldonado Garcia, 2014, 2015a), belongs to the Indo-Aryan language family and employs an extended Perso-Arabic script known as "Nastaliq" for writing (Ahmed, 2009; Maldonado Garcia, 2014). The diacritics within Urdu are recognized as the foundational components of the language's vowel system (Malik, 2010). Notably, Urdu encompasses eighteen vowels, comprising eight long oral vowels, seven long nasal vowels, three short vowels, and a multitude of diphthongs (Hussain, 2004; Saleem et al., 2002).

Figure 1. Urdu Vowel System and Orthographic Representation

Urdu letter	Roman Equivalents
Zabar	a
Zer	i
Paish	u
alif ا	a, aa
bari-ye ے	ai , ay, ei, e
chooti ye ی	ee, ey, i , ie
vao و	oo, au, ou, o, u

The diacritic system of Urdu is represented as follows:

Figure 2. Urdu Diacritic Marks



Source: Raza (2009)

Maldonado Garcia's works (2014, 2015a, 2015b) delve into the substantial influence of Arabic on Urdu, presenting extracted percentages that denote the prevalence of loanwords from Arabic within Urdu. Specifically, 21.7% of the Urdu lexicon is traced back to Persian borrowings, while an additional 29.9% originates from Arabic (Sabbour & Shafait, 2013; Memon et al., 2008). Despite Urdu's development from Sanskrit (Maldonado Garcia, 2018), the influence of Arabic remains significant, a point also acknowledged by scholars.

The connection between Urdu and the Nastaliq script is rooted in historical events. Following the Arab invasion of 712 AC, the Urdu language became associated with the Nastaliq script, which was crafted around the 1400s by Mir Ali Tabrizi (Maldonado Garcia, 2014; 2015; 2018; King, 2001). This script normalization aimed to align Urdu writing with religious texts inscribed in Arabic. In the Arabic writing system (Abjad), consonantal context holds prominence, with vowels indicated through marks above or below the preceding consonant. This practice is similar to Urdu, where speakers draw upon their existing knowledge to infer unwritten vowels (Raza & Hussain, 2010).

It's noteworthy that short vowels and similar diacritics are generally omitted in Arabic scripts (Zitouni, Sorensen & Sarikaya, 2006). Nonetheless, this omission presents challenges for readers, particularly those in the early stages of learning to read Urdu, second language Urdu learners, and individuals with disabilities. Thus, the absence of diacritics in Urdu reading materials can lead to confusion and hinder comprehension for these groups.

1.1 Objectives

Within this context, the study aims at achieving the following objectives:

- To ascertain the role of diacritics in facilitating comprehension of Urdu text contexts.
- To investigate the impact of diacritic markings on enhancing word identification, comprehension, and pronunciation skills among children reading Urdu.
- To evaluate the level of comprehension exhibited by 3rd and 4th grade students in Pakistani government schools when engaging with Urdu texts devoid of diacritics training.
- To gain insights into the challenges encountered by children while reading Urdu textbooks lacking diacritics.

1.2 Research Questions

1. What is the function of diacritics and understanding the context of the text produced in the Urdu language?
2. How do diacritic texts assist children in the improvement of fluency, comprehension, elocution, and word identification in the Urdu language?
3. How do students of the 3rd and 4th grades of government schools understand Urdu text without being trained in diacritics?
4. What kind of issues do children face in reading Urdu content without diacritics (their Urdu books)?

2. Literature Review

In Arabic, diacritics are typically not used by native speakers in their daily writing or reading, except in the case of religious texts like the Quran, where they aid non-native speakers in accurate pronunciation. Several studies advocate for the use of diacritics in Arabic, which in turn supports their implementation in Urdu. Abu-Rabia (1997) highlighted diacritics' significance in clarifying the homographic phenomenon, as they help in phonological disambiguation (Wiley et al., 2016). Perea et al. (2016) observed quicker processing of data with diacritics. Abdelhadi et al. (2011) found that 6th graders responded more accurately and rapidly to diacritic-inclusive texts compared to 3rd graders, who struggled with non-diacritic texts in reading Arabic. Hermena et al. (2015) affirmed the benefits of vowelized text in enhancing reader comprehension, particularly in cases where the lack of diacritics hindered Arabic comprehension.

Diacritics find usage in numerous languages. In the Malay context, where the phonological system entails 10 vowel sounds, contrasted with Arabic's three vowels—alif (ا), ya (ي), and wau (و)—a divergence is evident (Salahuddin & Winskel, 2014). The importance of Arabic in the Malay context arises due to the majority of Malaysians being Muslim and requiring Arabic reading proficiency for religious texts.

In Hebrew, the absence of diacritics hampers word identification, as indicated by Weiss, Katzir, and Biten (2016). Schiff (2003) conducted a study comparing vowelized and non-vowelized Hebrew texts among children, reporting positive outcomes favoring vowelized texts. He asserted that vowelized text forms the foundation for building initial reading skills and successfully interpreting un-vowelized scripts.

Diacritics in Romance languages serve to emphasize specific vowels, creating semantic nuances. For instance, the Spanish word "clave" means "password," while "clavé" transforms into the verb "clavar" in the first person, meaning "to nail, I nailed." This semantic distinction is widespread in Romance languages. Diacritic marks in Romance languages typically appear above vowels, such as á, é, í, ó, ú. The Spanish "dieresis," represented as "ü," and the "virgulilla," the sole diacritic above a consonant "ñ," further enrich Spanish phonetics with a distinct /ɳ/ pronunciation.

In French, there are 5 types, where one of them goes underneath the “c”:

1. ç – the cedilla
2. é – the acute accent
3. â/ê/î/ô/û – the circumflex
4. à/è/ì/ò/ù – the grave accent
5. ë/ï/ü – the trema

In the Greek language, diacritic marks signify stress positions (Daniels & Bright, 1996). Protopapas and Gerakaki (2009) pointed out that orthographic signs extracted from the lexicon in written form offer key insights into lexical stress patterns.

In Turkish, diacritics are referred to as accents and their absence presents issues related to homographic ambiguity. Arslan (2015) introduced a deASCIIfication approach to resolve this ambiguity. Ayçiçeği and Harris (2002) also noted that accented words are read more swiftly.

Similar to Urdu, Persian employs diacritics primarily for children and language learners (Mirdehghan, 2010). Vowel-free words in Persian can be phonologically opaque, but with diacritics indicating vowel forms, phonological clarity is achieved.

In Urdu, discussions about restoring diacritics often pertain to software applications. Javed and Hussain (2009) worked on developing an Urdu OCR for recognizing the Nastaliq Script. Raza and Hassan (2010) explored automatic diacritization for Urdu, highlighting its significance for computational tasks. Satti and Saleem (2012) addressed diacritics as complexities in developing an offline Urdu Nastaliq OCR. The absence of diacritics poses not just a reading challenge but also computational hindrances.

Children learning to read face a primary issue due to the absence of diacritics. This concern is less pronounced for adult readers in languages like Spanish, Urdu, and Arabic, as they can deduce pronunciation and meaning from context. In Arabic, studies like Abdelhadi et al. (2011), Hermena et al. (2015), and Perea et al. (2016) have shown that diacritics significantly influence reading, with vowelized texts aiding comprehension.

Regarding Urdu, limited research exists. Farrukh and Vulchanova (2014) examined reading accuracy and speed in children, touching on diacritics but not extensively. The complexity of Urdu's orthography was noted. This study aims at addressing the gap by focusing on children's reading difficulties due to the absence of diacritics, a facet often overlooked in the literature.

3. Methodology

To investigate the challenges children, encounter while reading texts without diacritics in Urdu, a mixed methods approach has been employed. This study encompasses both quantitative and qualitative methodologies to provide a comprehensive understanding of the issue.

The quantitative aspect of the research employs an experimental design, utilizing a sample acquired through a simple random sampling technique. This sample is subsequently divided into two distinct groups: the experimental group, exposed to the variable under scrutiny, and the control group, which does not experience this variable.

Ellis (1994) introduces a comprehensive framework for understanding learners' errors, underlining their significance and the factors that influence them. Central to this model is the systematic collection of a well-defined sample from learner language, enabling the clear delineation of the types of errors learners make while considering contextual conditions. Ellis' model draws heavily from Corder's error analysis model (1981), encompassing five stages for error analysis:

- Collection: A sample of data is systematically collected for the purpose of analysis.
- Identification: Errors within the collected data are identified and isolated.
- Description: The identified errors are meticulously described and categorized.
- Explanation: An explanation is offered to elucidate the reasons behind the occurrence of each error.
- Evaluation: A critical evaluation of potential strategies to address and rectify the identified errors is undertaken.

By adopting this analytical model, the study aims at gaining a comprehensive insight into the issues children encounter while reading Urdu texts without diacritics. The combination of quantitative experimental data and the qualitative analysis of errors provides a robust foundation for understanding and addressing these challenges.

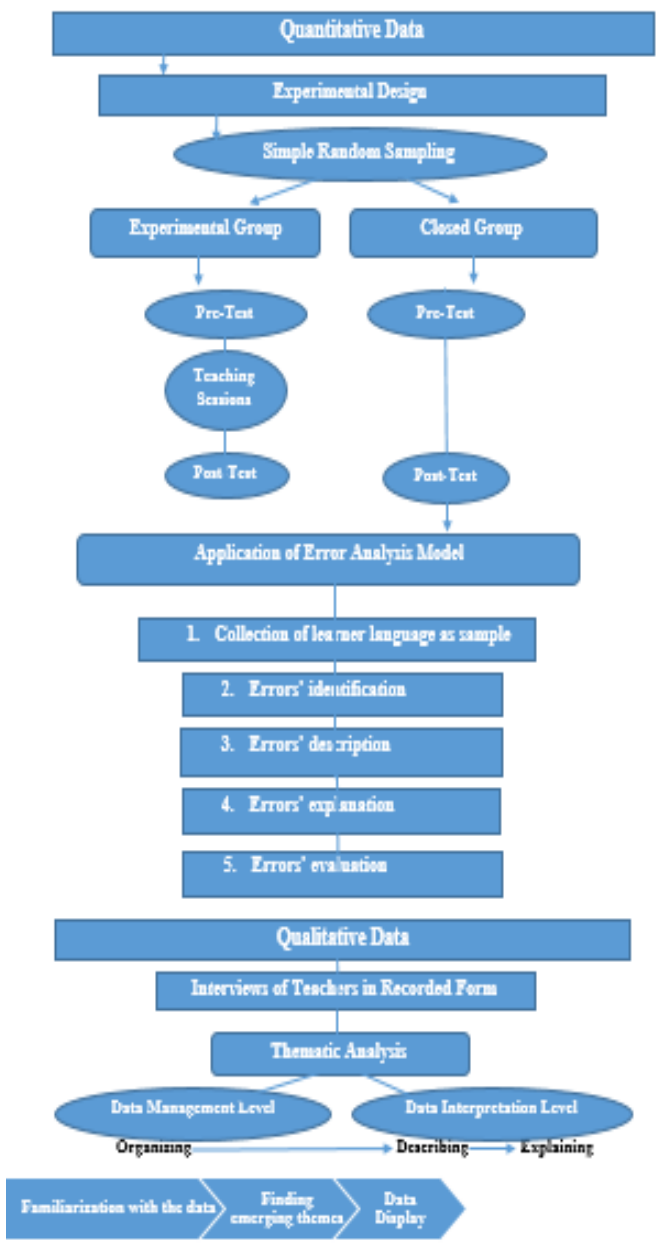
3.1 Sample

A cohort of 40 students from grades 3 and 4, drawn from two government schools in (XX City), was chosen as the sample using a simple random technique, from a population of 60 students. The selection of 3rd and 4th graders was purposeful, as these students, aged around 9 and 10 years, have typically developed foundational reading skills. In 3rd grade, reading is primarily utilized for learning purposes, while by 4th grade, students employ reading to gain diverse perspectives on various subjects.

The participants chosen exhibited proficient reading skills without encountering significant errors. Their visual acuity was normal, and they did not present any language disorders. These grades were specifically targeted due to the assumption that students at this stage should have attained competence in both vowelized and non-vowelized reading. Additionally, their reading activities encompass both learning and acquiring multiple viewpoints.

The participants' native language was Punjabi, with Urdu serving as their second language. Notably, Urdu serves as both the medium of instruction in Pakistani schools and their second language of instruction.

Figure 3. Structure of Methodology



3.2 Materials

The corpus employed for the tests was derived from the children's textbooks. To formulate the pre and post-tests, a list comprising both homographic and non-homographic words was compiled from the respondents' textbooks. Subsequently, these words were integrated into the pre-test, which was administered to both groups. The pre-test consisted of text without diacritics, which participants read while errors were recorded.

The theoretical framework devised by Ellis was subsequently applied. Initially, 40 individual pre-test trials were conducted in a soundproof environment, with 25 trials conducted in each of the schools (A and B). Following the pre-test, the experimental group underwent a period of reading instruction involving diacritics. During this instructional phase, students were introduced to the concept of diacritics as vowel markers. This was followed by reading exercises focusing on homographic words with diacritics. Students were instructed to read words with diacritics in the manner they were taught.

Subsequently, the post-test was administered to the experimental group after the teaching phase. The post-test, similar to the pre-test, involved 50 individual trials conducted within soundproof rooms (25 in each school). In total, 100 experimental trials were conducted throughout the study.

3.3 Interviews

To ensure triangulation and enhance the reliability of the findings, semi-structured interviews were conducted with 20 teachers from grades 2, 3, and 4. These interviews were aimed at gaining diverse perspectives and cross-validating the results obtained. The interviews were meticulously designed to offer a balanced yet flexible framework for discussion.

Both audio and video recordings were made during the interviews, capturing not only the spoken words but also the nuances of non-verbal communication and gestures. Subsequently, the recordings were transcribed, ensuring a detailed textual representation of the interviews' content.

This multifaceted approach to data collection, including both student trials and teacher insights, contributes to a comprehensive understanding of the issues at hand and augments the credibility of the study's outcomes.

3.4 Limitations of the Study

The study's scope is centered around schools within (XX City), without encompassing educational institutions from other provinces of Pakistan. Moreover, it exclusively evaluates the reading proficiency of children in grades 3 and 4, without differentiating by gender. Additionally, the study's focus is confined to the assessment of four specific diacritics ("Zer," "Zabar," "Pesh," and "Tashdeed"), chosen due to their prevalence in high-frequency words featured in the selected grade-level textbooks.

It is worth noting that future research endeavors could extend the model and methodology to encompass diverse geographical regions within Pakistan. This expansion could provide a more comprehensive perspective on the impact of diacritics on reading skills across the country, considering variations in curricula, linguistic backgrounds, and regional influences.

3.5 Ethical Considerations

The study adhered to ethical guidelines and ensured proper informed consent protocols. Written informed consent was obtained from the parents of all participating students, as the respondents were minors. Furthermore, the school principals also granted informed consent, permitting the research to be conducted within their educational institutions. The teachers involved in the study were also provided with detailed information and gave their informed consent to participate.

It's important to highlight that the research procedures were conducted in accordance with the principles outlined in the Declaration of Helsinki, which offers ethical guidelines for studies involving human participants. This underscores the commitment to upholding the well-being and rights of all individuals involved in the research process.

4. Analysis of the Results

Diacritics play a crucial role in the process of learning to read Urdu. The Urdu language employs the Nastaliq script, a Perso-Arabic script, where the absence of diacritic marks poses challenges for children as the vowels remain unmarked. Introducing diacritic marks to indicate vowels significantly enhances children's reading fluency by addressing potential impairments and minimizing errors.

This study aligns with Ellis's model of error analysis (1994, 2005), which comprises five essential components for a comprehensive error analysis. These elements were systematically applied throughout this study, ensuring a rigorous examination of the subject matter.

4.1 Collection and Identification of Errors

The pre-tests and post-tests were both recorded in audio and video formats, subsequently transcribed for analysis. During this phase, errors were collected and identified, forming a crucial step in the process. Specifically, a total of 31 errors were gathered and identified from the pre-test assessments.

Table 1. Phonological Errors of the Students (Pre-test)

Sr. No.	Correct Pronunciation	Urdu Script	English Meaning	Error of Pronunciation	Urdu Script	English Meaning	Frequency
1	/hʊsn/	حُسن	Beauty	/həsən/	حَسَن	Arabic Name (good-looking)	5
2	/həsən/	حَسَن	Arabic Name (good-looking)	/hʊsn/	حُسن	Beauty	2
3	/ələm/	عَلَم	Black flag	/ɪlm/	عِلم	Knowledge	9
4	/ɪlm/	عِلم	Knowledge	/ələm/	عَلَم	Black flag	3
5	/ɑxɪr/	آخِر	At last / Finally	/ɑxər/	آخر	Untranslatable	5
6	/ʊn/	اُن	Far (Used to refer people and objects)	/ɪn/	اِن	Near (Used to refer people and objects)	3
7	/ɪn/	اِن	Near (Used to refer people and objects)	/ʊn/	اُن	Far (Used to refer people and objects)	6
8	/ɪɖʰr/	اِدھر	Here	/ʊɖʰr/	اُدھر	There	5
9	/ʊɖʰr/	اُدھر	There	/ɪɖʰr/	اِدھر	Here	5
10	/ʧʰotɪ/	چھوٹی	Little	/ʧʰʊtɪ/	چُھوٹی	Left	2
11	/ʧʰʊtɪ/	چُھوٹی	Left	/ʧʰotɪ/	چھوٹی	Little	2
12	/ɪnʰẽ/	اِنہیں	Near (Used to refer people and objects)	/ʊnʰẽ/	اُنہیں	Far (Used to refer people and objects)	4
13	/ʊnʰẽ/	اُنہیں	Far (Used to refer people and objects)	/ɪnʰẽ/	اِنہیں	Near (Used to refer people and objects)	5
14	/qəsəm/	قَسَم	Swear	/qɪsəm/	قِسم	Kind/Type	7
15	/qɪsəm/	قِسم	Kind/Type	/qəsəm/	قَسَم	Swear	1
16	/mæl/	مَیل	Dirt	/meɪl/	میل	Meet	6

				/mɪːl/	مِیل	Mile	2
17	/mɪːl/	مِیل	Mile	/meɪl/	میل	Meet	8
18	/ɪs/	اِس	This	/ʊs/	اُس	That	1
19	/ʊs/	اُس	That	/ɪs/	اِس	This	11
20	/kʊʧʰ/	کُچھ	Some	/kəʧʰ/	کچھ	Underarm (In Punjabi)	5
21	/kɪj'ja/	کِیا	Done	/kɪja/	کیا	What	3
22	/kɪja/	کیا	What	/kɪj'ja/	کِیا	Done	3
23	/ʈ̪ʔlʊq/	تعلُق	Relation	/ʈ̪ʔləq/	تعلق	Untranslatable	2
24	/ɪse/	اِسے	Near (Used to refer people and objects)	/ʊse/	اُسے	Far (Used to refer people and objects)	3
25	/ʊse/	اُسے	Far (Used to refer people and objects)	/ɪse/	اِسے	Near (Used to refer people and objects)	4
26	/kɪsɪ/	کِسی	Anyone	/kəsɪ/	کسی	Untranslatable	1
27	/kʊɾsi/	کُرسی	Chair	/kəɾsi/	کرسی	Untranslatable	1
28	/ʊməɾ/	عُمر	Age / Urdu Name	/əməɾ/	عمر	Untranslatable	2
29	/ɗɪn/	دِن	Day	/ɗən/	دن	Untranslatable	3
30	/ʤʊmlõ/	جُملوں	Sentences	/ʤəmlõ/	جملوں	Untranslatable	1
31	/bəʧe/	بچے	Remaining	/bəʧ'ʧe/	بچّے	Children	8
Total Number of Errors							**128**

The presented table illustrates two versions of a single word functioning as homographs, each possessing distinct meanings. During the assessment, a number of students read one version as the other and continued to do so for each respective version. Among these errors, some are categorized as inter-lingual, while others are classified as intra-lingual. The students faced challenges in distinguishing between the two versions in the absence of short vowel marks. All these errors are identified as phonological errors resulting from the lack of diacritics. As displayed in Table 1, a total of 31 errors were identified, exhibiting a range of frequencies spanning from a maximum of eleven occurrences to just one occurrence.

In the post-test phase, certain students from the experimental group exhibited the repetition of the same errors, which can be viewed as mistakes. Despite their diligent practice, these repeated errors appear to have become ingrained in their language usage—a phenomenon known as fossilization. However, in some instances, students were able to self-correct. On the other hand, several students displayed positive responses to the new concepts introduced during the teaching sessions, correctly applying the taught principles.

4.2 Description of Errors

Ellis (1994) posits those errors "arise when the learner creates a deviant structure on the basis of other structures in the target language." However, it's crucial to acknowledge that in Pakistan, children are not instructed in reading using the Punjabi language. Consequently, the participants in this study were solely engaged in reading, and not writing. Consequently, the errors observed predominantly possessed a phonological nature. These errors were closely linked to the pronunciation of words, emerging when participants struggled to accurately identify specific words.

Furthermore, the errors primarily surfaced in situations where respondents encountered difficulties in distinguishing between homographs. Even when considering the context of the text, participants were often unable to differentiate between the various parts of speech, further contributing to the errors in their reading.

4.3 Explanation of Errors

The participants' native language was Punjabi, while their primary medium of instruction and target language was Urdu. Both languages share the same Nastaliq script, yet in the Punjab region, children are typically not educated in reading and writing the Punjabi language, focusing solely on spoken communication. Urdu serves as their medium of instruction, being the national and official language of Pakistan.

The errors that were collected predominantly pertained to intra-lingual issues, stemming from two main factors. Firstly, the respondents lacked a comprehensive understanding of the rules governing the target language (L2), which is Urdu in this context. Secondly, the school curriculum did not encompass specific instructions and lesson plans pertaining to diacritics. Instead, over time, students tended to grasp diacritics through self-learning, context, and reinforcement. This absence of formal instruction on diacritics contributed to the occurrence of these errors.

4.4 Error Evaluation

While certain studies may overlook this step in the application of Error Analysis, it is indeed crucial to ascertain the significance of individual errors. Diacritic errors are prevalent across numerous languages. Given the linguistic intricacies of the Urdu language encompassing morphology and syntax, these phonological errors can have broader implications. In order to quantify and assess the severity of these errors, an endeavor was undertaken to translate a subset of errors. The outcomes of this translation exercise reveal that these errors often lead to substantial deviations in terminology, indicating incongruities within the context of the reading process. In certain instances, the errors even convey a meaning diametrically opposite to that of the original term. This latter scenario denotes a particularly grave error, magnifying its impact on comprehension.

4.5 Pre-test and Post-Test Comparison

The comparison between the pre-test and post-test holds significant value as it sheds light on the origins of errors and their categorization as either inter-lingual or intra-lingual. This comparative analysis effectively addresses the first research question. Furthermore, this comparison serves to underscore the advancement and growth observed in students subsequent to the diacritics teaching intervention.

As indicated in Table 1, the pre-test exhibited a total of 31 errors arising from non-diacritic text. However, in the post-test phase following the teaching sessions, the number of errors reduced considerably to just 13 errors. This reduction in errors serves as a tangible testament to the positive impact of the diacritics instruction on the students' reading proficiency.

Table 2. Phonological Errors of the Students (post-test)

Sr. No.	Correct Pronunciation	Urdu Script	English Meaning	Error of Pronunciation	Urdu Script	English Meaning	Frequency
1	/hʊsn/	حُسن	Beauty	/həsən/	حَسَن	Arabic Name (good-looking)	1
2	/ələm/	علَم	Black flag	/ɪlm/	عِلم	Knowledge	5
3	/ɑxɪɾ/	آخِر	At last / Finally	/ɑxəɾ/	آخر	Untranslatable	2
4	/ʊn/	اُن	Far (Used to refer people and objects)	/ɪn/	اِن	Near (Used to refer people and objects)	1
5	/ɪnʰẽ/	اِنہیں	Near (Used to refer people and objects)	/ʊnʰẽ/	اُنہیں	Far (Used to refer people and objects)	2
6	/qɪsəm/	قِسم	Kind/Type	/qəsəm/	قَسَم	Swear	1
7	/mæl/	مَیل	Dirt	/meɪl/	میل	Meet	3
8	/mɪːl/	مِیل	Mile	/meɪl/	میل	Meet	4
9	/mæl/	مَیل	Dirt	/mɪːl/	مِیل	Mile	1
10	/ʊs/	اُس	That	/ɪs/	اِس	This	1
11	/kɪj'ja/	کِیا	Done	/kɪja/	کیا	What	1
12	/kʊɾsi/	کُرسی	Chair	/kəɾsi/	کرسی	Untranslatable	1
13	/bəʧe/	بَچے	Remaining	/bəʧ'ʧe/	بچّے	Children	3
Total Number of Errors							**26**

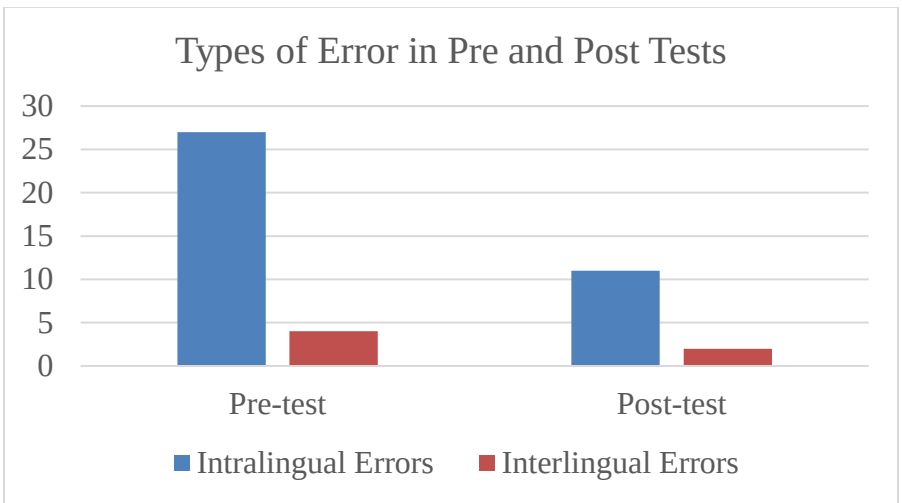
Table 3 illustrates the various types of errors committed by the students, along with their corresponding quantities.

Table 3. Phonological Errors of the Students (pre-test vs post-test)

Tests	Intralingual Errors	Interlingual Errors	Errors (Words)	Total number of Errors
Pre-test	27	4	31	128
Post-test	11	2	13	26

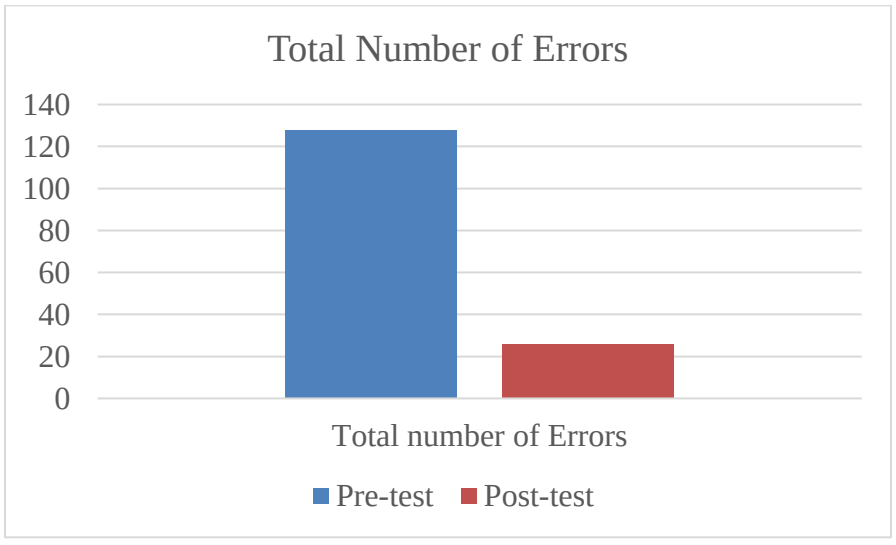
Graph 1 visually represents the contrast in the types of errors made by the students. In the pre-test phase, a total of 27 intra-lingual errors were recorded, accompanied by 4 inter-lingual errors. However, in the post-test phase, the number of intra-lingual errors dropped to 11, while inter-lingual errors reduced to just 2. This graphical illustration visually underscores the change in error patterns between the pre-test and post-test assessments:

Graph 1. Comparison in the types of the errors of the students



Graph 2 visually showcases the disparity in error frequencies observed between the pre-test and post-test phases. This graphical representation unequivocally illustrates the efficacy of diacritics instruction. It substantiates the premise that such instruction is instrumental in aiding students to differentiate between homographic words and develop effective reading skills.

Graph 2. Comparison in the frequencies of the errors of the students



Graph 2 effectively demonstrates a significant reduction in error frequency following the implementation of instructions.

In terms of qualitative data analysis, the researchers engaged in interviews with 20 teachers affiliated with the respondents. Notably, it's worth mentioning that out of the entire teacher sample, only 2 were Urdu teachers. Nevertheless, the study ensured the application of a triangulation approach, a technique aimed at enhancing both the validity and reliability of the findings. The interviews were conducted in Urdu, and the responses were subsequently translated into English to serve the research purposes.

The teachers were presented with the following questions during the interview process:

4.6 Interview Questions

1. Do children face problems while reading Urdu text?
2. What kind of problems and difficulties do they face in reading?
3. How do children read homographic words used without diacritics?
4. Do they take more time in recognizing those words used without diacritics than all other words in the textbook or lesson?
5. What role do diacritics play in reading Urdu words?
6. Do you find any importance, if Urdu text/words are introduced with diacritics? / Do you see any importance of using diacritics?
7. Do they play any role in understanding the context of the lesson, in word recognition and in pronunciation?
8. In which grade do children face more problems and why?
9. What strategies or steps do you take to ensure the children read correctly when they make errors related to diacritics?
10. What do you think if the text with diacritics is introduced instead of plain text, would it be more helpful in disambiguating Urdu texts for children? Would it have any importance or not? /Should it be introduced by the government and till which grade they must be used?

Indeed, despite the researchers already having a positive answer to question 1, which inquired, "Do children face problems while reading Urdu text?", this question was deliberately included in the teacher interviews. Its inclusion aimed to establish a correlation between the outcomes of the pretest and the observations and experiences of the teachers. Remarkably, the responses obtained from the teachers in response to this question were unanimous. They uniformly confirmed that students encounter challenges while reading texts devoid of diacritics. This alignment between the teacher insights and the pretest results underscores the significance and accuracy of the findings.

Question 2, which inquired about the types of problems students encounter while reading, prompted teachers to provide insights into the primary errors students make during the reading process. The teachers' responses unveiled that the predominant errors students encounter include the following:

1. Errors in the pronunciation of words
2. Errors in word identification
3. Errors of context understanding

All the teachers unanimously concurred that the respondents indeed make the three types of errors outlined. Additionally, certain teachers highlighted other errors, which encompass challenges such as:

1. Difficulty in reading new words
2. Struggles with connecting letters to form complete words
3. Pronunciation hurdles while encountering complex words
4. Identifying specific sounds

Moving on to questions 3 and 4, which inquired, "How do children read homographic words used without diacritics?" and "Do they take more time in recognizing those words used without diacritics than all other words in the textbook or lesson?", the teachers' responses conveyed that students exhibit pauses linked to identification when dealing with homographic words. These pauses, lasting a fraction of a second, are utilized to differentiate such words. Moreover, the teachers indicated that students encounter additional challenges:

1. Difficulty in reading unfamiliar words
2. Trouble joining letters to form words
3. Errors arising from missing certain letters within intricate words
4. Instances of omitting specific sounds while reading words
5. Occurrence of inter-lingual errors

The teachers' insights echo these sentiments, underlining the multifaceted nature of the challenges faced by students when confronted with homographic words and diacritic-absent text.

> “Yes, they get stuck on the word if they do not know the pronunciation while trying to recognize the word. In this process, they take 5-6 mini seconds pause.”

> “They can’t identify the word. So they do not pronounce fully as they can’t identify…Mistakes related to diacritics, they also do occur due to lack of usage and knowledge of them like /kɪsmət/ will be read as /kəsmət/.”

In addressing questions 5, 6, and 7, which queried, "What role do diacritics play in reading Urdu words?", "Do you find any importance if Urdu text/words are introduced with diacritics? / Do you see any importance of using diacritics?", and "Do they play any role in understanding the context of the lesson, in word recognition, and in pronunciation?", the teachers expressed unanimous support for the inclusion of diacritics in students' textbooks.

The teachers unanimously endorsed the significance of incorporating diacritics into Urdu text. They affirmed that diacritics hold importance in aiding students' understanding of context within lessons, facilitating word recognition, and enhancing pronunciation accuracy. The teachers' unified stance underscores the essential role diacritics play in optimizing students' reading experience and language acquisition.

> “They are the backbone of Urdu language as they tell about the sound system of Urdu.....They are crucial.”

> “Without diacritics, Urdu language is incomplete for, reading as they tell about the vowel system of Urdu language. Without these diacritics, basic concepts of the homographic words can’t be taught.”

The answers of the teachers reflect the following important points:

1. All the teachers favored its use for providing clarity in reading.
2. The majority of errors are found to be intra-lingual and this point emphasizes on teaching and using diacritics.
3. Teachers’ agreement on the need of writing diacritics and their suggestions regarding the implementation of diacritics in the books until children develop the habit of reading with diacritics, is another point laying stress on using short marks.
4. Teachers tell that it would be beneficial for children if they read with diacritics after learning a particular concept or term.

Addressing question 8, which inquired about the grade in which children encounter more reading problems and the reasons behind it, the teachers' responses reflected varying perspectives. While some teachers highlighted the 2nd grade as a pivotal stage, others pointed towards both the 2nd and 3rd grades. Their reasons included insufficient practice at home, the substantial leap from 1st to 2nd and 3rd grades, the introduction of a new reading pattern not encountered in the 1st grade, and the progression to a slightly more advanced level where the

complexity of reading materials increases. The teachers also noted that many students might not possess the cognitive capacity to grasp the concept of homographs. Moreover, the absence of diacritic marks emerged as a technical reason amplifying the difficulties encountered.

Turning to question 9, which inquired about strategies or steps adopted to rectify errors related to diacritics, the teachers acknowledged their efforts to provide students with strategies that encompass diacritics application and understanding words devoid of diacritics through contextual cues. However, the notable limitation was that the teachers involved in this study were specialists in other subjects, not Urdu as a second language (L2). Consequently, they encountered challenges in teaching vowel markers since Urdu was not their area of expertise. While some teachers suggested activities like board exercises, dedicated reading sessions, and focused instruction on homographic words, no specific activities were identified to teach the phonological nuances of the Urdu language. It was revealed that only two teachers among the entire sample were Urdu specialists. Moreover, the teachers' own awareness of this phonological phenomenon appeared limited. Given that the majority of teachers were not Urdu teachers, but specialists in other subjects, they encountered difficulties in addressing this issue effectively.

Lastly, addressing question 10, which examined the teachers' perspectives on the introduction of diacritic-marked text versus plain text, the responses exhibited a range of viewpoints. A significant portion of teachers favored the implementation of diacritics, emphasizing their efficacy in enhancing various aspects of reading. These teachers believed that diacritics indeed aid in clarifying the meaning of Urdu text for children. However, it's worth noting that some teachers held differing opinions on this matter, suggesting a diversity of viewpoints within the teacher sample.

> "Students will pick plain or simple text more quickly as compared to the text with diacritics because they practiced with the plain text".

Without diacritics training, children tend to overlook vowel marks or become stuck. However, following the training, diacritics were recognized as facilitators that enhance fluency, text comprehension, pronunciation, and word recognition. This observation further underscores that the chosen textbooks inconsistently incorporate diacritics on the necessary words, with their placement appearing random. The mixed methods approach definitively demonstrated that children commonly encounter difficulties when reading texts lacking diacritics. Consequently, the study underscores the potential for substantial improvements in their reading experiences and skills through a few adjustments to their textbooks.

5. Conclusion

Urdu, a language significantly influenced by Arabic (Maldonado Garcia, 2015b), contrasts with Arabic where diacritics are not essential for word recognition due to Arabic speakers' adeptness in reading without them. This study involved a comparison of text reading with and without diacritics in the Urdu language, focusing on primary school children. Employing a mixed methods approach, the analysis unveiled a remarkable enhancement in word recognition when diacritics were introduced. The pre-test identified 128 errors, which markedly diminished to 26 errors in the post-test after diacritics were instructed. It's noteworthy that all the teachers strongly endorsed diacritics instruction and advocated for the inclusion of diacritics lessons in textbooks to enhance children's reading capabilities.

Funding: This study was not funded in any shape or form by any party.

Conflict of Interest: The authors declare that they have no conflict of interest.

Bio-note:

Maria Isabel Maldonado Garcia holds a PhD in General Linguistics from UNED, Spain. Recognized for her outstanding contributions, she has been honored with the prestigious Fatima Jinnah National Pride Award. Currently serving as the Incharge and Professor at the Institute of Languages and Linguistics, University of the Punjab, Maria Isabel's expertise and dedication continue to enrich the field of linguistics and education. Her remarkable journey and significant achievements make her a trailblazer in her domain, inspiring generations to come.

Muzamal Haider is an accomplished educator and scholar in the realm of English Language and Literature. Currently serving as a Lecturer in the esteemed Department of English Language and Literature at the University of Lahore, located in the vibrant province of Punjab, Pakistan, Muzamal's academic journey has been marked by dedication, enthusiasm, and a commitment to fostering intellectual growth.

References

Abdelhadi, S. I. R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research*, *3*(2), 117-127.

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing: An Interdisciplinary Journal, 9*, 65–78.

Ahmed, T. (2009) Roman to Urdu transliteration using wordlist. In *Proceedings of the Conference on Language and Technology*, *305,* 309.

Al-Samarraie, H., Sarsam, S. M., Alzahrani, A. I. (2020) Reading text with and without diacritics alters brain activation: The case of Arabic. *Curr Psychol 39*, 1189–1198. https://doi.org/10.1007/s12144-019-00493-6

Arslan, A. (2016). Deasciification approach to handle diacritics in Turkish information retrieval. *Information Processing & Management: an International Journal, 52*(2) 326-339. Retrieved from: https://doi.org/10.1016/j.ipm.2015.08.004

Ayçiçeği, A., & Harris, C. L. (2002). How are letters containing diacritics represented? Repetition blindness for Turkish words, *European Journal of Cognitive Psychology*, *14*(3), 371-382. DOI: https://doi.org/10.1080/09541440143000113

Corder, S. P. (1981), *Error Analysis and Interlanguage*. Oxford: Oxford University Press. 5-13.

Dai, J., Ibrahim, R., & Share, D. L. (2013). The influence of orthographic structure on printed word learning in Arabic. *Writing Systems Research*, *5*(2), 189-213.

Daniels, P. T., & Bright, W. (Eds.). (1996). *The World's Writing Systems.* Oxford: Oxford University Press. 3-17.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 494–507. https://doi.org/10.1037/xhp0000032

Hussain, S. (2004). Letter to sound rules for Urdu text to speech system. In *Proceedings of Workshop on Computational Approaches to Arabic Script-based Languages*. Geneva, Switzerland: Coling. 74-79.

Javed, S. T., & Hussain, S. (2009). Improving Nastalique specific pre-recognition process for Urdu. In *Proceedings (IEEE) 13th International Multitopic Conference*, Islamabad.1-6. Doi: https://doi.org/10.1109/INMIC.2009.5383111

Kurzon, D. (2013): Diacritics and the Perso-Arabic script, *Writing Systems Research*, *5*(2), 234-243 Doi: 10.1080/17586801.2013.799451

King, R. D. (2001). The poisonous potency of script: Hindi and Urdu. *International journal of the sociology of language, 2001*(150), 43-59.

Maldonado García, M. I. (2014). The Urdu Language Reforms. *Almas 15*, 14-24.

Maldonado García, M. I. (2015a). Reforming the National Curriculum. Teaching the National and Provincial Languages of Pakistan. *Journal of Elementary Education*, *25*(2), 73-87.

Maldonado Garcia, M. I. (2015b). A Corpus Based Quantitative Survey of the Persian and Arabic Elements in the Basic Vocabulary of Urdu Language. *Pakistan Vision, 16*(1), 63-94.

Maldonado Garcia, M. I. (2018). A Corpus Based Quantitative Survey of The Etymological Elements in The Basic Vocabulary of Urdu Language. *Almas Urdu Research Journal, 20*(1) XI-XXVI.

Malik, M. G. A. (2010). *Methods and Tools for Weak Problems of Translation*. *Computer Science*. Doctoral Dissertation. University Joseph-Fourier – Grenoble.

Maqbool, S. N., Ahmed, G., Saleem, M. M., & Iqbal, (2015) An Efficient Hindi-Urdu Transliteration System. *Science International, 27*(4), 4549-4553.

Memon, J. A., Khowaja, K., & Kazi, H. (2008). Evaluation of steganography for Urdu/Arabic text. *Journal of Theoretical & Applied Information Technology, 4*(3).

Mirdehghan, M. (2010). Persian, Urdu, and Pashto: A comparative orthographic analysis. *Writing Systems Research*, *2*(1), 9-23.

Perea, M., Abu Mallouh, R., Mohammed, A., Khalifa, B., & Carreiras, M. (2016). Do diacritical marks play a role at the early stages of word recognition in Arabic? *Frontiers in Psychology*, *7*, 1255. Doi: https://doi.org/10.3389/fpsyg.2016.01255

Protopapas, A., & Gerakaki, S. (2009). Development of processing stress diacritics in reading Greek. *Scientific Studies of Reading*, *13,* 453–483. Doi: https://doi.org10.1080/10888 430903034788

Raza, A., & Hussain, S. (2010) Automatic Diacritization for Urdu. In *Proceedings of the Conference on Language and Technology*. 105-111.

Sabbour, N., & Shafait, F. (2013). A segmentation-free approach to Arabic and Urdu OCR. In *Document recognition and retrieval XX* (8658)215-226.

Saleem, A. M., Kabir, H., Riaz, M. K., Rafique, M. M., Khalid, N., & Shahid, S. R. (2002). Urdu consonantal and vocalic sounds. Center for Research in Urdu Language Processing. CRULP Annual Student Report.

Salehuddin, K., & Winskel, H. (2014). Reducing the Cognitive Complexity in Reading the Arabic Script of Written Malay via Diacritics. *Procedia-Social and Behavioral Sciences, 114*, 40-44.

Satti, D. A., & Saleem, K. (2012). Complexities and implementation challenges in offline Urdu Nastaliq OCR. In *Proceedings of the Conference on Language & Technology* (CLT12). University of Engineering and Technology. 85–91.

Schiff, R. (2003). The effects of morphology and word length on the reading of Hebrew nominals. *Reading and Writing: An Interdisciplinary Journal, 16*, 263–287.

Weiss, Y., Katzir, T., & Bitan, T. (2016). When transparency is opaque: Effects of diacritic marks and vowel letters on dyslexic Hebrew readers. *Cortex, 83*, 145–159. Doi: https://doi.org/10.1016/ j.cortex.2016.07.017

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal Experimental Psychology: Human Perception,42*, 1186–1203.Doi: https://doi.org/10. 10 37/xhp0000213

Zitouni, I, Sorensen, J., & Sarikaya, R. (2006) Maximum Entropy Based Restoration of Arabic Diacritics. *Proceedings of the 21st International Conference on Computational Linguistics.* Sydney, Australia. Doi: https://10.3115/1220175.1220248

Appendix

Annex (A)

Pre and Post Test

Without Diacritics (As Pre-Test)	With Diacritics (As Post-Test)
اس کی بات سن کر میرا جسم سن ہو گیا	اُس کی بات سُن کر میرا جِسم سُن ہو گیا
حسن کو پھولوں کا حسن بہت پسند ہے	حَسَن کو پھولوں کا حُسن بہُت پسند ہے
علم علم کے بارے میں بتاتا ہے	عِلم علَم کے بارے میں بتاتا ہے
آخر اس نے علم کے ساتھ کیا کیا	آخِر اُس نے علَم کے ساتھ کیا کِیا
ان کا ان سے کوئی تعلق نہیں ہے	اُن کا اِن سے کوئی تعلُّق نہیں ہے
کسی سے ادھر ادھر کا کچھ مت کہنا	کِسی سے اِدھر اُدھر کا کُچھ مَت کہنا
چھوٹی کے ہاتھ سے رسی چھوٹی	چھوٹی کے ہاتھ سے رَسی چُھوٹی
انہیں منع کیا تھا کہ انہیں نہ بتاؤیں	اُنہیں منع کِیا تھا کہ اِنہیں نہ بتاؤیں
قسم اس خدا کی جس نے طرح طرح کی قسم کے پھل بنائے ۔	قَسَم اُس خُدا کی جِس نے طرح طرح کی قِسم کے پھل بنائے
عمر کے بچے نے بچے ہوئے آم کھائے	عُمر کے بچّے نے بچّے ہوئے آم کھائے
ایک میل چل کر میرے پاؤں پر میل جم گیا	ایک مِیل چل کر میرے پاؤں پر مَیل جم گیا
تم اسکی اور اسکی کتاب واپس دو	تُم اِسکی اور اُسکی کتاب واپس دو
جملوں کو غور سے سنو	جُملوں کو غور سے سُنو
آج کے دن سورج کا گرم ہونا ممکن ہے	آج کے دِن سُورج کا گرم ہونا مُمکن ہے
اسے کہو کرسی اسے دے دے	اُسے کہو کُرسی اِسے دے دے
اتنا بڑا جن نظر آیا اور چلا گیا	اِتنا بڑا جِن نظر آیا اور چلا گیا
مسلمان دن میں پانچ نمازیں پڑھتے ہیں	مُسلمان دِن میں پانچ نمازیں پڑھتے ہیں

Annex (B)

Interview Questions

1. Do children face problems while reading Urdu text?
2. What kind of problems and difficulties do they face in reading?
3. How do children read homographic words used without diacritics?
4. Do they take more time in recognizing those words used without diacritics than all other words in the textbook or lesson?
5. What role do diacritics play in reading Urdu words?
6. Do you find any importance, if Urdu text/words are introduced with diacritics? / Do you see any importance of using diacritics?
7. Do they play any role in understanding the context of the lesson, in word recognition and in pronunciation?
8. In which grade do children face more problems and why?
9. What strategies or steps do you take to ensure the children read correctly when they make errors related to diacritics?
10. What do you think if the text with diacritics is introduced instead of plain text, would it be more helpful in disambiguating Urdu texts for children? Would it have any importance or not? /Should it be introduced by the government and till which grade they must be used?