




Evaluating the Readability of English Instructional Materials in Pakistani Universities: A Deep Learning and Statistical Approach



Muhammad Saqlain 

Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, 10140 Bangkok, Thailand

* Correspondence: Muhammad Saqlain (muhammad.saql@kmutt.ac.th)

Received: 07-21-2023

Revised: 09-10-2023

Accepted: 09-15-2023

Citation: Saqlain, M. (2023). Evaluating the readability of English instructional materials in Pakistani universities: A deep learning and statistical approach. *Educ. Sci. Manag.*, 1(2), 101-110. <https://doi.org/10.56578/esm010204>.



© 2023 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: In educational settings of Pakistan, where English is utilized as the primary medium of instruction but not as an official language, the assessment of instructional text readability is crucial. This research investigates the impact of text readability on student comprehension and achievement by integrating deep learning methods with mathematical and statistical approaches. It has been observed that when suitably trained, deep learning models exhibit a significant correlation with human assessments of text readability. The investigation further illuminates the linguistic and structural elements influencing readability. Such insights are instrumental for educators and content developers in establishing standards to craft more accessible educational materials. Emphasis is placed on the exploration of Advanced Natural Language Processing (NLP) techniques, the incorporation of multilingual models, and the refinement of curricular structures to enhance readability assessments. Additionally, the study underscores the necessity of engaging with educational policymakers in Pakistan to implement accessibility guidelines. These efforts aim to reduce linguistic barriers, amplify student potential, and foster an inclusive educational ecosystem. The findings and methodologies presented in this study offer a comprehensive understanding of the challenges and solutions in optimizing English language instructional materials for non-native speakers, with potential applications in diverse multilingual educational contexts.

Keywords: English as language; Statistics; Deep learning; Higher education system; Text readability; Natural Language Processing (NLP)

1. Introduction

In the landscape of higher education in Pakistan, where English is not an official language, the assessment of instructional text readability in colleges and universities emerges as a crucial aspect. With the predominance of English as the medium of instruction, it becomes imperative to ensure the clarity and accessibility of these materials for students from diverse linguistic backgrounds. This investigation explores the impact of instructional text readability on students' comprehension and academic achievement, addressing the challenges posed by the use of English as a teaching medium. By tailoring instructional materials to the linguistic diversity of the student body, educational institutions can bridge the language gap, offering an equitable and productive learning environment for all students, irrespective of their proficiency in English.

Recognized as the de facto global lingua franca, English's role in facilitating personal and professional development among college students is undeniable. With over 2 billion speakers and official status in 45 countries, its significance in international communication and knowledge exchange in science, technology, and culture is paramount. In Pakistan, English proficiency is increasingly becoming a vital skill in the global talent market, enabling cultural integration and assimilation of global innovations. Stemming from the colonial legacy of British rule, English is frequently employed as the medium of instruction in Pakistani higher education institutions, particularly in disciplines like science, technology, and business (Carfax Educational Projects, 2015). Consequently, the assessment of English language proficiency has become a critical component of the higher education system in Pakistan, where students are often expected to demonstrate fluency in the language.

Standardized tests such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) are commonly used to evaluate students' reading, writing, speaking, and listening skills (Shamim, 2011). However, the reliance on English as the primary instruction language and these standardized assessments can create educational access barriers, especially for students from socioeconomically disadvantaged backgrounds and regions where English is less prevalent. The prerequisite of English competence for admissions may restrict opportunities for numerous aspiring students (Murray, 2019; Shen, 2018).

The use of standardized language assessments in Pakistan's higher education system has not been without its critics. Questions have been raised regarding their validity, with some scholars arguing that these tests may not accurately reflect academic or practical language proficiency (Asif et al., 2020; Canale & Swain, 1980). Furthermore, a socioeconomic divide is evident, as students who can afford test preparation materials generally score higher on these exams, exacerbating educational opportunity disparities (Imran, 2011). In response, some Pakistani universities are exploring alternative methods for evaluating language proficiency, such as written assignments and interviews, to provide a more comprehensive representation of students' abilities (Eng et al., 2013). Additionally, preparatory courses and language support programs are being developed to enhance students' English proficiency, both prior to and during their university studies. Ultimately, assessing students' English language proficiency in the context of its use as a medium of instruction remains a fundamental aspect of higher education in Pakistan (Haruna et al., 2019; Ma, 2021). Despite the significance of standardized testing, efforts to enhance the inclusivity and accessibility of the system continue, acknowledging the importance of equitable access and a more thorough assessment of language proficiency (Saqlain, 2023).

Numerous students in Colleges and Universities (CAUs) encounter substantial challenges in mastering English, as evidenced by difficulties in improving their English Proficiency Level (EPL). The complexity of the curriculum is often cited as a significant barrier to their progress. In response, academic research has delved into diverse instructional methodologies to enhance College English Teaching (CET) (Abid & Saqlain, 2023; Haq & Saqlain, 2023b). Findings suggest that game-based vocabulary education has yielded effective results, enhancing motivation and engagement among learners (Cortes-Robles et al., 2021; Haq & Saqlain, 2023a; Lee et al., 2021; Liu & Tsai, 2021; Saqlain et al., 2020). Moreover, a stratified approach to CET has been associated with more successful teaching outcomes (Huang, et al., 2020; Song & Chen, 2021; Sekmen et al., 2021; Saqlain & Xin, 2020). Despite these advancements in the field of Public English research and pedagogical innovation within CAUs, a notable gap persists in the research concerning English reading materials (Maulana & Sanusi, 2021).

Deep learning, utilized in conjunction with statistical and mathematical methodologies, has been employed to address this gap. The results of this study indicate a promising avenue for tackling the challenges of readability in educational texts. It has been established that deep learning models, when properly trained and calibrated, are capable of assessing and analyzing the readability of instructional texts. These models demonstrate a strong correlation with human evaluations of readability, suggesting their potential as reliable tools in determining the suitability of educational resources. Furthermore, the application of statistical and mathematical methods in this study provides insights into the linguistic and structural elements of texts that significantly influence readability. Future research may focus on the impact of curriculum modifications on the holistic educational experiences of students. The advancement of educational accessibility could be furthered by developing and refining deep learning models for assessing the readability of educational materials in various languages, including local languages. Such endeavors would not only enhance the inclusiveness of the educational system but also potentially contribute to a more comprehensive understanding of the factors influencing language acquisition and proficiency in multilingual educational settings.

2. Methodology and Mathematical Background

Deep learning has emerged as a transformative approach to addressing the challenges associated with evaluating English language competency in Pakistani colleges and universities. This methodology surpasses the limitations of traditional multiple-choice examinations. Automated Essay Scoring (AES), enabled by Advanced Natural Language Processing (NLP) models, facilitates a comprehensive assessment of writing skills, as detailed in studies (Hassan & Janjua, 2015; Khattak, 2012; Malik et al., 2020; Mahmood et al., 2021). Moreover, the application of deep learning in speech recognition enhances the thorough evaluation of speaking abilities, taking into account factors such as fluency and pronunciation. This technology enables more accurate and authentic testing of oral communication skills (Rehmani, 2003). Additionally, the integration of deep learning into the development of personalized language learning systems ensures tailored learning experiences, catering to individual proficiencies and challenges. Such advancements hold the potential to dismantle barriers to educational access in Pakistan, fostering a more equitable and efficient system for assessing and enhancing English language proficiency.

In the realm of text Automatic Readability Assessment (ARA), the versatility of Convolutional Neural Networks (CNN) positions them as highly effective tools for classification tasks (Ji et al., 2021). The methodology involves constructing source and target matrices $X_{1:T}$ and $Y_{1:T}$ from the corresponding language sequences $x_1; x_2; \dots; x_T$ and $y_1; y_2; \dots; y_T$, where each consists of word embedding vectors $x_t, y_t \in \mathbb{R}^k$ with a dimension of k . The

source matrix $X_{1:T}$ utilizes a convolution window $w_j \in \mathbb{R}^{l \times k}$ of length l , facilitating the convolution process to generate a series of feature maps. The Rectified Linear Unit (ReLU), known for its rapid convergence and straightforward gradient computation, is selected as the activation function. The process of pooling involves extracting the maximum value from the feature map $\tilde{c}_j = \max\{c_{j1}, c_{j2}, \dots, c_{jT-l+1}\}$ to obtain the final features pertinent to w_j . This output is then flattened through a fully connected layer, integrating various components: h representing neuron weight, X as the input, σ as the sigmoid activation function, and \tanh for the hyperbolic tangent activation function (Havaei et al., 2021).

Long Short-Term Memory (LSTM) networks enhance the core module by incorporating a gating mechanism, consisting of three sigmoid neural network layers, combined with element-wise multiplication. This improvement facilitates superior management of information flow (Umma et al., 2020). The \tanh activation function is primarily utilized for managing data within the state and output operations (Lyashevskaya et al., 2021). In the LSTM architecture, the input gate i_t , forget gate f_t , cell state C_t , output gate o_t , and unit output h_t at time t play pivotal roles. The input gate governs the incorporation of upper unit output information into the current layer's unit information while preserving historical data. The calculation of each threshold is meticulously detailed, ensuring accuracy and efficiency in the process.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

In the outlined Eqs. (1)-(3), the threshold weight, denoted by W , and its corresponding offset, symbolized by b , are critical factors (Haq & Saqlain, 2023b). Subsequent to each threshold's update, the cell state C_t undergoes a revision, as encapsulated in Eq. (4).

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

where, W_c and b_c represent the current weight and offset, respectively. The output gate plays a pivotal role in controlling the neural network output weight of the cell state. The activated unit state is then transmitted to the subsequent layer of the neural network and the chain unit (Haq & Saqlain, 2023a).

$$h_t = o_t * \tanh(C_t) \quad (5)$$

Eq. (5) delineates this process, where σ is identified as a sigmoid activation function. This function is responsible for outputting the network's memory state and compressing sequential input data into the $[0,1]$ range. In this framework, a zero value signifies the non-transfer of information, while a value of one indicates the complete transfer of information (Dobres, et al., 2017; Gao, et al., 2020; Perkins, 2019; Zhu et al., 2018). The output is derived by multiplying the matrix by the computation's result at the current layer, which is then inputted into the next layer, contingent upon the result adhering to a specified range. Eq. (6) provides the mathematical representation of this function, mapping the real number field to the $[0,1]$ interval and indicating the probability of the data belonging to the positive class.

$$f(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

Differing from the sigmoid, the hyperbolic tangent activation function (\tanh) maps the real number field to the $[-1,1]$ range, outputting zero when the input is zero (Liu & Tsai, 2021). The expression for the tanh activation function is presented in Eq. (7).

$$f(x)_{\tanh} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

In the field of NLP, the contextual representation of words is influenced by both preceding and subsequent information. The Bi-LSTM (BiLSTM) network, which incorporates both a forward and a backward LSTM, enhances sensitivity to the context surrounding a word. This feature distinguishes Bi-LSTM from standard LSTM

(Zhang et al., 2020; Mahmoud & Zrigui, 2021; Bausch et al., 2021). The present study investigates the application of English text ARA as a model training process to learn from a dataset labeled with text levels. Here, the labeled text dataset is denoted as $\{(d_1, l_1), (d_2, l_2), \dots, (d_n, l_n)\}$, D represents the set of labeled texts, n the total number of labeled texts, d_i the i -th labeled text in D , and $l_i \in \{G_1, G_2, \dots, G_m\}$ the corresponding readability level of the text in this dataset, m signifies the number of distinct readability levels (Denise, et al., 2000).

3. Calculation

3.1 Actor-Resource-Activity (ARA) Model

The ARA model's proficiency in extracting key concepts for gauging text readability from a corpus of tagged texts is a focal aspect of its functionality. This model operates within a framework comprising these extracted concepts (Qin & Zeng, 2018; Zhai et al., 2022; Wang & Feng, 2021). Notably, the model exhibits adaptability to new textual resources, demonstrating its capability to effectively ascertain the readability of texts beyond those in the initial dataset (Wang et al., 2021). The principal objective of the deep learning-based English text ARA model is the mapping of the text dataset $D = \{d_1, d_2, \dots, d_n\}$ to its associated readability level $l_i = \{G_1, G_2, \dots, G_m\}$. The model is structured into three components: text representation, feature extraction, and text categorization, as depicted in Figure 1.

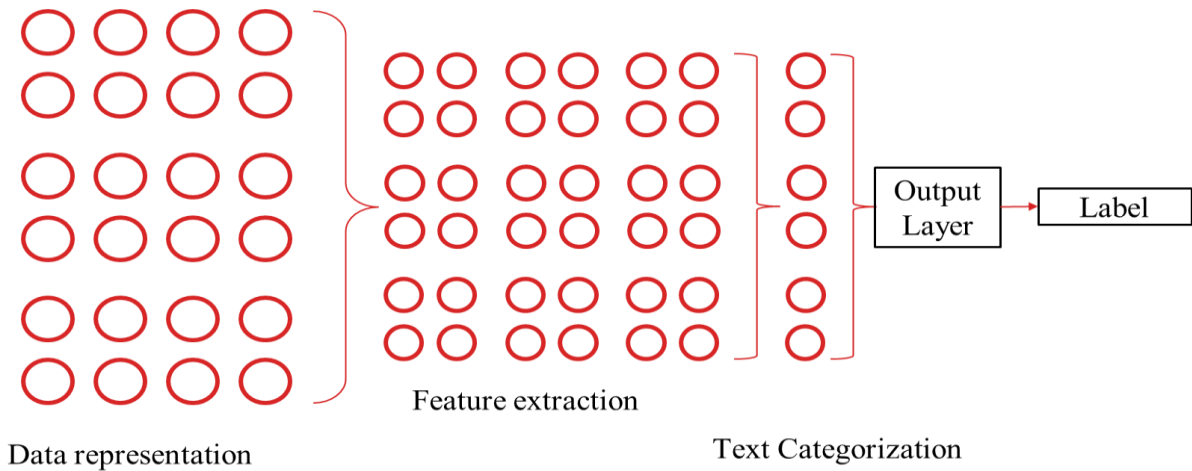


Figure 1. Structure of the proposed deep learning-based ARA model

The application of convolution kernels of varying sizes facilitates the discernment of word relationships across different ranges. One convolution kernel is set to a specific size k . As this kernel traverses the text, it encounters a window matrix at each position in the sequence, denoted as i . This matrix $\bar{W}_i = [x_i, x_{i+1}, \dots, x_{i+k-1}]$, comprising consecutive words, undergoes convolution with the kernel matrix to yield a characteristic graph (Weldemariam et al., 2020). At position i , the feature mapping of the word window vector is computed as per Eq. (8), where \otimes represents multiplication, and b signifies the offset. The sigmoid function (σ) is employed in this calculation.

$$c_i = \sigma(w \otimes m + b) \quad (8)$$

In the text classification phase, a multi-classifier based on logistic regression is utilized. This classifier processes the final representation vector retrieved from the pooling layer and inputs it into the softmax layer for categorization, employing the feature vector generated by the CNN. Eq. (9) delineates the precise calculation for this process.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}} \quad (9)$$

where, z_i signifies the output of the i -th node. The constant C represents both the total number of output nodes and the classification categories assigned to each node. The softmax function transforms a multi-class input into a probability distribution within the $[0,1]$ range, ensuring the summation of the probabilities equals 1. The cross-entropy loss function is selected for the English text-oriented ARA model, with its calculation detailed in Eq. (10).

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (10)$$

where, M denotes the number of classes. The variable y_{ic} is defined such that if the model-predicted class aligns with the sample i 's class, y_{ic} equals 1; otherwise, it is 0. p_{ic} represents the estimated probability that the observed sample i belongs to category c . The dataset construction process is outlined in Figure 2, utilizing textbooks of varying grades to represent different difficulty levels.

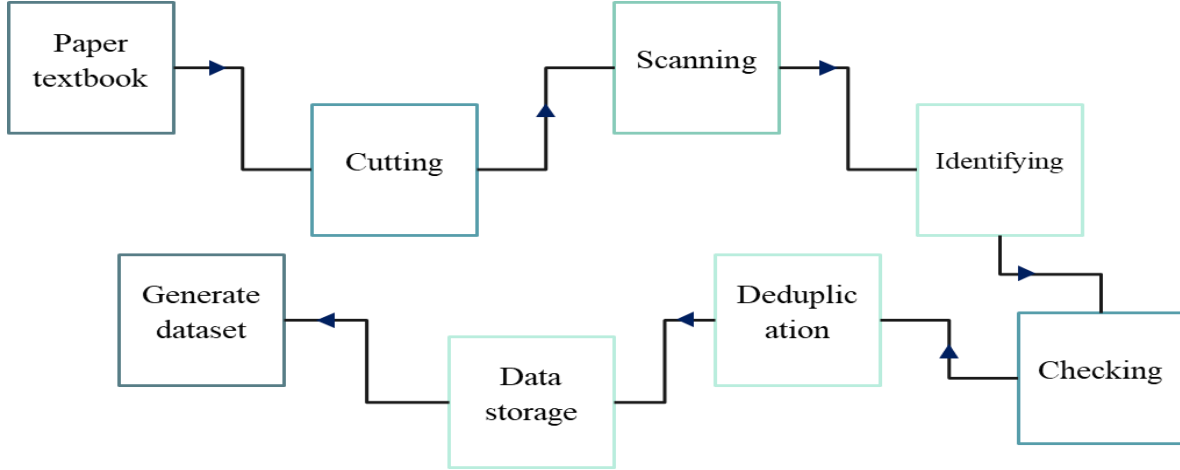


Figure 2. Dataset construction process

The digitization process of the textbook content is depicted in Figure 2. Initially, textbooks are segmented into individual pages, which are then rapidly scanned to produce images using a high-speed scanner. The extracted text undergoes manual verification via an optical character recognition interface. During this verification, teachers filter out content not matching targeted difficulty levels, and researchers correct any misidentified language. Duplicate entries are removed using Excel's data deduplication tool, culminating in the data storage phase. Through this methodology, a dataset comprising 1,000 phrases of varied difficulty is compiled, categorized into levels 1-4. For the English text-oriented ARA model, 30% of this dataset is randomly selected as the test set, with the remaining 70% allocated for training. The training parameters include a batch size of 64, a dropout rate of 0.5, and kernel sizes of 3, 4, and 5, with each size contributing 100 feature maps, resulting in a total of 200 feature maps. The learning rate for the model is set at 0.001.

3.2 Statistical Analysis

Statistical methods significantly enhance the evaluation of English language competency within Pakistani colleges and universities. By analyzing large datasets, these methods identify patterns and trends in students' language proficiency, thus providing a comprehensive understanding of their strengths and weaknesses. Item Response Theory (IRT) exemplifies a statistical technique that augments assessment accuracy by considering both the difficulty level of each question and the individual student's ability, resulting in a more precise proficiency estimation. Additionally, the validity and reliability of existing tests are scrutinized using statistical methods to ensure they accurately represent the targeted language proficiency. Furthermore, statistical analyses are instrumental in identifying socioeconomic and demographic factors influencing language competency, thereby facilitating targeted interventions to bridge educational access disparities. The application of statistical methods in the assessment process culminates in a fairer and more individualized evaluation of English language competency in the context of Pakistani higher education.

This study focuses on first-year students at Sargodha University, comprising two classes: Class A (the experimental group) and Class B (the control group), each with 25 students. These students were selected based on comparable College Entrance Examination (CEE) English scores, indicating a similar baseline in English proficiency. The university randomly assigned students to their respective classes. A preliminary analysis of the pre-test data revealed no significant differences in the EPL for reading between the two groups. The distribution and demographics of the test subjects were thoroughly examined, revealing a concentration of student scores in the 75-80 points range in both classes. The mean scores of Class A and Class B were closely aligned, averaging 71.65 and 71.24, respectively. The continuous variation in scores within each class indicated that the EPL in

reading was essentially equivalent for both groups. The experimental protocol is divided into distinct stages. Stage 1 involves administering identical reading assessments to both groups to establish a baseline. The ARA model's predictions align with the students' assessments of text difficulty. An independent sample t-test is conducted to validate the parallel status of the two groups, confirming the absence of significant differences. In the subsequent stage, reading materials are curated and categorized into four distinct levels (1-4) based on the reading EPL of the experimental group, employing the proposed deep learning-based ARA technique. While the control group receives materials at random, irrespective of their reading level, the experimental group is provided with materials one level above their current reading EPL.

The final stage involves analyzing the gathered data using Statistical Package for the Social Sciences (SPSS) and other analytical tools. IRT (Embretson & Reise, 2013) is also applied to explore the relationship between individuals' performances on a test item and the test takers' levels of performance on an overall measure of the ability that the item was designed to measure.

4. Findings and Result Discussion

The effectiveness of the model is evaluated through its accuracy in classifying the entire sample set. The model outputs readability scores for individual sentences, and the overall text readability is determined by aggregating these sentence-level scores. This aggregation involves calculating the text's maximum difficulty level, which is achieved by weighting and summing the quantity of sentences across various difficulty levels. The pre-test text was segmented into four distinct sections, each assigned a difficulty rating from 1 to 4 based on the proposed ARA model. However, these sections were classified into three levels of difficulty according to the "English Reading Grading Guidance": basic for the first level, intermediate for the second and third levels, and advanced for the fourth. This juxtaposition illustrates that the ARA model's assessment is more granular compared to the national standard, thereby offering a more nuanced evaluation of the readability of college-level English texts.

The preliminary test results for Class A and Class B were analyzed, with a focus on the accuracy of the students' responses to the reading materials. In this analysis, the pre-test reading scores serve as the test variable, and class designation as the grouping variable. An independent t-test was employed to evaluate the data. The EPL reading proficiency of Class A and Class B showed no statistically significant difference, with the pre-test reading scores yielding a significance level of 0.848, surpassing the 0.05 threshold. This finding substantiates the equivalence of the two classes, qualifying them as "parallel classes." The pre-test results were further categorized into four levels of difficulty: scores 0-25 were labeled as difficulty 1, 26-50 as difficulty 2, 51-75 as difficulty 3, and 76-100 as difficulty 4. Based on these categories, the reading proficiency assessments of the students in Class A and Class B are presented in Figure 3.

Subsequent to the teaching experiment, both Class A and Class B underwent a post-test assessment. The proportion of students achieving full reading comprehension at each difficulty level was calculated. Figure 4 details these post-test results.

The statistical analysis of the post-test results is depicted in Figure 5.

For the analysis, 'class' was designated as the grouping variable, and 'pre-test reading score' as the test variable. An independent t-test revealed a post-test reading proficiency significance of 0.006, which falls below the 0.05 threshold, indicating a significant difference between the reading EPL of Class A and Class B. This result underscores the enhancement of college students' reading proficiency through the implementation of the deep learning-based ARA model tailored for English texts.

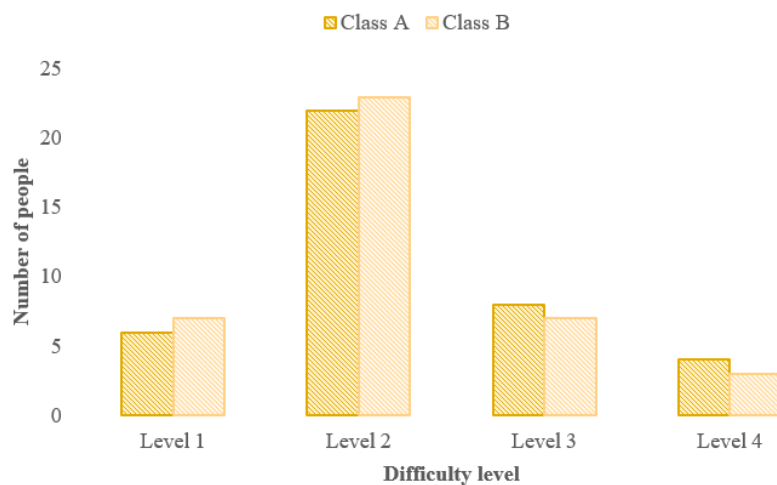


Figure 3. Grading of Class A and Class B



Figure 4. Post-test results of Class A and Class B

1	Class	Covariance	Confidence level	Error range	Lower confidence limit	Upper confidence limit
2	Class A	51.42	0.95	2.50	68.62	65.86
3	Class B	41.76	0.95	2.25	73.61	70.37

Figure 5. Post-test statistics of Class A and Class B

The post-test results were further categorized into four levels of difficulty: scores 0-25 were classified as difficulty 1, 26-50 as difficulty 2, 51-75 as difficulty 3, and 76-100 as difficulty 4. Figure 5 delineates the distribution of reading proficiency levels for students in Classes A and B based on these criteria. A comparative analysis of Class A's pre- and post-test results reveals a notable improvement in the students' English reading EPL. After the intervention, an increase was observed in Class A, where two additional students achieved comprehension at difficulty level 2, and eleven students advanced from level 2 to level 3 in reading proficiency. Conversely, in Class B, only a modest improvement was noted: two students progressed from level 2 to level 3, and five students improved from reading difficulty level 1 to level 2.

These findings validate the effectiveness of the newly developed ARA model for college English, leveraging deep learning techniques. While standardized tests have traditionally played a crucial role, there is a growing recognition of the need for accuracy, fairness, and inclusivity in assessing students' language proficiency. In Pakistan, efforts are being made to address the challenges related to English language evaluation. Academic institutions are increasingly exploring alternative methods for evaluating linguistic fluency, such as interviews and written tasks, which offer a more comprehensive view of learners' abilities. Initiatives are also underway to provide preparatory courses and language support to students from diverse linguistic backgrounds. Furthermore, the study has yielded valuable insights into the linguistic and structural elements of texts that influence readability. Utilizing this information, guidelines for educators and content developers can be formulated to facilitate the creation of instructional materials that are more accessible and user-friendly, accommodating a broad spectrum of students.

5. Conclusion

This study has delved into the implications of using English as the medium of instruction in Pakistani colleges and universities, underscoring the necessity of assessing students' English language proficiency as a critical component of higher education. By examining the impact of text readability on student comprehension and academic success, the research addressed challenges associated with English as the instructional medium. The utilization of statistical and mathematical algorithms, in conjunction with deep learning techniques, facilitated the evaluation of instructional texts' readability. The findings suggest that these methodologies are effective in assessing the suitability of instructional materials, evidenced by their capacity to analyze text readability and their significant correlation with human assessments.

As higher education institutions evolve and refine their language assessment strategies, it is imperative to consider both the student impact and the overarching goal of ensuring equitable access to quality education. The requirement of English language proficiency for admission to Pakistan's premier colleges and universities poses a challenge for students struggling with English, particularly affecting individuals from non-English speaking backgrounds. This issue potentially restricts their educational and employment opportunities.

6. Future Directions

Collaborative efforts between educators and content developers could lead to curriculum adaptations that more effectively accommodate students' linguistic diversity. The potential development of multilingual deep learning models, capable of evaluating text readability in various languages, including regional dialects, presents an exciting avenue for future research. Advanced Natural Language Processing techniques, such as sentiment analysis and content summarization, hold promise for enhancing instructional materials and augmenting student comprehension and engagement. By actively soliciting feedback from educators and students, it will be possible to tailor algorithms and models to align with the evolving needs of the educational community.

Data Availability

The authors will supply the supporting information upon request, free from any unjustified limitations, which forms the foundation of the conclusions presented in this article.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- Abid, M. & Saqlain, M. (2023). Decision-making for the bakery product transportation using linear programming. *Spectr. Eng. Manag. Sci.*, 1(1), 1-12. <https://doi.org/10.31181/sems1120235a>.
- Asif, S., Afzal, I., & Bashir, R. (2020). An analysis of medium of instruction policies in the education system of Pakistan with specific reference to English medium education. *Sir Syed J. of Educ. Social Res.*, 3(2), 370-382. [https://doi.org/10.36902/sjesr-vol3-iss2-2020\(370-382\)](https://doi.org/10.36902/sjesr-vol3-iss2-2020(370-382)).
- Bausch, J., Subramanian, S., & Piddock, S. (2021). A quantum search decoder for natural language processing. *Quantum Mach. Intell.*, 3(1), 16. <https://doi.org/10.1007/s42484-021-00041-1>.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Appl. Ling.*, 1(1), 1-47. http://www.uefap.com/tefsp/bibliog/canale_swain.pdf
- Carfax Educational Projects (2015). *English Language in Pakistan Higher Education*. London: British Council.
- Cortes-Robles, O., Barocio, E., Obushevs, A., Korba, P., & Sevilla, F. R. S. (2021). Fast-training feedforward neural network for multi-scale power quality monitoring in power systems with distributed generation sources. *Meas.*, 170, 108690. <https://doi.org/10.1016/j.measurement.2020.108690>.
- Denise, A., Roques, O., Termier, M. (2000). Random generation of words of context-free languages according to the frequencies of letters. *Math. Comput. Sci.*, 113-125. https://doi.org/10.1007/978-3-0348-8405-1_10.
- Dobres, J., Chrysler, S. T., Wolfe, B., Chahine, N., & Reimer, B. (2017). Empirical assessment of the legibility of the highway gothic and Clearview signage fonts. *Transp. Res. Rec.*, 2624(1), 1-8. <https://doi.org/10.3141/2624-01>
- Embretson, S. E. & Reise, S. P. (2013). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Eng, L. S., Mohamed, A. R., & Javed, M. (2013). Analysis of students' competency in listening comprehension of the English language at Pakistani secondary school level. *Middle-East J. Sci. Res.*, 16(3), 331-341. <http://doi.org/10.5829/idosi.mejrs.2013.16.03.2298>.

- Gao, R., Tang, Y. C., Xu, K. W., Huo, Y. K., Bao, S. X., Antic, S. L., Epstein, E. S., Deppen, S., Paulson, A. B., Sandler, K. L., Massion, P. P. & Landman, B. A. (2020). Time-distanced gates in long short-term memory networks. *Med. Image Anal.*, 65, 101785. <https://doi.org/10.1016/j.media.2020.101785>.
- Haq, H. B. & Saqlain, M. (2023a). An implementation of effective machine learning approaches to perform Sybil Attack Detection (SAD) in IoT network. *Theor. Appl. Comput. Intell.*, 1(1), 1-14. <https://doi.org/10.31181/taci1120232>.
- Haq, H. B. U. & Saqlain, M. (2023b). Iris detection for attendance monitoring in educational institutes amidst a pandemic: A machine learning approach. *J. Ind. Intell.*, 1(3), 136-147. <https://doi.org/10.56578/jii010301>.
- Haruna, H., Zainuddin, Z., Mellecker, R. R., Chu, S. K., & Hu, X. (2019). An iterative process for developing digital gamified sexual health education for adolescent students in low-tech settings. *Inf. Learn. Sci.*, 120(11/12), 723-742. <https://doi.org/10.1108/ILS-07-2019-0066>.
- Hassan, S. & Janjua, F. (2015). A comparative study of English assessment in Pakistani Context. *Kashmir J. of Lang. Res.*, 18(3), 203-220.
- Havaei, N., Mahmoodzadeh, M., Azad, A., Meimandi, M., Rezaei, M., Kahjoogh, M. A., & Mohammadi, A. (2021). Handwriting status among Iranian primary school students: A cross-sectional study. *Int. J. Therapy Rehabil.*, 28(2), 1-9. <https://doi.org/10.12968/ijtr.2020.0007>.
- Huang, J. H., Di, X. G., Wu, J. D., & Chen, A. Y. (2020). A novel convolutional neural network method for crowd counting. *Front Inf. Technol. Electron. Eng.*, 21(8), 1150-1160. <https://doi.org/10.1631/FITEE.1900282>.
- Imran, K. H. A. N. (2011). Reading assessment techniques among selected secondary school teachers in Pakistan: Current trends and practices. *Int. J. New Trends Educ.*, 2(4), 58-141. <https://www.ajindex.com/dosyalar/makale/acarindex-1423904123.pdf>
- Ji, F., Zhang, H., Zhu, Z., & Dai, W. (2021). Blog text quality assessment using a 3D CNN-based statistical framework. *Future Gener. Comput. Syst.*, 116, 365-370. <https://doi.org/10.1016/j.future.2020.10.025>.
- Khattak, S. G. (2012). Assessment in schools in Pakistan. *SA-eDUC*, 9(2). https://neqmap.bangkok.unesco.org/wp-content/uploads/2019/08/6_Assessment-in-schools-in-Pakistan.pdf
- Lee, C. C., Rahiman, M. H. F., Rahim, R. A., & Saad, F. S. A. (2021). A deep feedforward neural network model for image prediction. *J. Phys. Conf. Ser.* 1878(1), 012062. <https://doi.org/10.1088/1742-6596/1878/1/012062>.
- Liu, L. & Tsai, S. B. (2021). Intelligent recognition and teaching of English fuzzy texts based on fuzzy computing and big data. *Wirel. Commun. Mob. Comput.*, 2021, 1170622. <https://doi.org/10.1155/2021/1170622>.
- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49, 100529. <https://doi.org/10.1016/j.asw.2021.100529>.
- Ma, Y. (2021). The application of schema theory in the teaching of English reading in senior high schools. *Region-Educational Res. Rev.*, 3(3), 17-20. <https://doi.org/10.32629/rerr.v3i3.412>.
- Mahmood, S., Hussain, H. B., & Mahmood, S. (2021). Reading assessment awareness: A comparative analysis of public and private school ESL teachers in Pakistan. 5(1), 12-22, [http://doi.org/10.47205/plhr.2021\(5-1\)1.2](http://doi.org/10.47205/plhr.2021(5-1)1.2).
- Mahmoud, A. & Zrigui, M. (2021). BLSTM-API: Bi-LSTM recurrent neural network-based approach for Arabic paraphrase identification. *Arabian J. Sci. Eng.*, 46, 4163-4174. <https://doi.org/10.1007/s13369-020-05320-w>.
- Malik, S., Khan, A., & Sadiq, U. (2020). A comparative analysis of assessment schemes in Secondary School Certificate and Cambridge O Level English examination papers in Pakistan: Need for reform. *J. Humanities Social Sci.*, 28(1), 47-70. <http://ojs.uop.edu.pk/jhss/article/view/385>
- Maulana, D. & Sanusi, A. (2021). The clarity and legibility of students' translation of Arabic - Indonesian official text. *J. Ilmiah Program Studi Pendidikan Bahasa Arab Iain Palangka Raya*, 9(1), 49-66. <https://doi.org/10.23971/altarib.v9i1.2659>.
- Murray, H. (2019). More than 2 billion pairs of eyeballs: Why aren't you sharing medical knowledge on Wikipedia?. *BMJ Evidence-Based Med.*, 24(3), 90-91. <http://dx.doi.org/10.1136/bmjebm-2018-111040>.
- Perkins R. A. (2019). Assessment and evaluation in online learning. *Lib. Technol. Rep.*, 55(4), 31-34.
- Qin, Y. & Zeng, Y. F. (2018). Research of clinical named entity recognition based on Bi-LSTM-CRF. *J. Shanghai Jiaotong University (Sci.)*, 23, 392-397. <https://doi.org/10.1007/s12204-018-1954-5>.
- Rehmani, A. (2003). Impact of public examination system on teaching and learning in Pakistan. *Int. Biannual Newsletter Antriep*, 8(2), 3-7.
- Saqlain, M. & Xin, X. L. (2020). Interval valued, m-polar and m-polar interval valued neutrosophic hypersoft sets. *Neutrosophic Sets and Sys.*, 36. <https://doi.org/10.5281/zenodo.4065475>.
- Saqlain, M. (2023). Sustainable hydrogen production: A decision-making approach using VIKOR and Intuitionistic Hypersoft Sets. *J. Intell. Manage. Decis.*, 2(3), 130-138. <https://doi.org/10.56578/jimd020303>.
- Saqlain, M., Jafar, N., Moin, S., Saeed, M., & Broumi, S. (2020). Single and multi-valued neutrosophic hypersoft set and tangent similarity measure of single valued neutrosophic hypersoft sets. *Neutrosophic Sets and Syst.*, 32(1), 317-329. <https://doi.org/10.5281/zenodo.3723165>.
- Sekmen, A., Parlaktuna, M., Abdul-Malek, A., Erdemir, E., & Koku, A. B. (2021). Robust feature space separation for deep convolutional neural network training. *Discover Artif. Intell.*, 1, 1-11. <https://doi.org/10.1007/s44163-021-00013-1>.

- Shamim, F. (2011). *English As the Language for Development in Pakistan: Issues, Challenges and Possible Solutions*. London: British Counsel. pp. 291-310.
- Shen, G. R. (2018). Chinese college English teachers' ability to develop students' informationized learning in the era of big data: Status and suggestions. *Eurasia J. of Math., Sci. Technol. Educ.*, 14(6), 2719-2728. <https://doi.org/10.29333/ejmste/90554>.
- Song, Y. & Chen, C. (2021). MPPCANet: A feedforward learning strategy for few-shot image classification. *Pattern Recogn.*, 113, 107792. <https://doi.org/10.1016/j.patcog.2020.107792>.
- Umma, R. J., Adisaputera, A., & Pramuniati, I. (2020). Development of text review assessment model using authentic assessment in 11th grade students of senior high school Negeri 1 Dolok Batunanggar, Simalungun Regency. *Budapest Int. Res. Critics in Ling. Educ. J.*, 3(2), 1099-1112. <https://api.semanticscholar.org/CorpusID:226083187>
- Wang, P., Li, X. D., & Wu, R. L. (2021). A deep learning-based quality assessment model of collaboratively edited documents: A case study of Wikipedia. *J. Inf. Sci.*, 47(2), 176-191. <https://doi.org/10.1177/0165551519877646>.
- Wang, W. K. & Feng, A. (2021). Self-information loss compensation learning for machine-generated text detection. *Math. Prob. in Eng.*, 2021, 6669468. <https://doi.org/10.1155/2021/6669468>.
- Weldemariam, D. G., Amaha, N. D., Abdu, N., & Tesfamariam, E. H. (2020). Assessment of completeness and legibility of handwritten prescriptions in six community chain pharmacies of Asmara, Eritrea: A cross-sectional study. *BMC Health Serv. Res.*, 20, 1-7. <https://doi.org/10.1186/s12913-020-05418-9>.
- Zhai, Y. J., Wan, Y., & Wang, X. X. (2022). Optimization of traffic congestion management in smart cities under bidirectional long and short-term memory model. *J. Adv. Transp.*, 2022, 3305400. <https://doi.org/10.1155/2022/3305400>.
- Zhang, W., Qin, J., Mei, F., Fu, J., Dai, B., & Yu, W. (2020). Short-term power load forecasting using integrated methods based on long short-term memory. *Sci. China Technol. Sci.*, 63(4), 614-624. <https://doi.org/10.1007/s11431-019-9547-4>.
- Zhu, W., Yao, T., Ni, J., Wei, B., & Lu, Z. (2018). Dependency-based Siamese long short-term memory network for learning sentence representations. *PloS One*, 13(3), e0193919. <https://doi.org/10.1371/journal.pone.0193919>.