# Assessing the readability of open data portals: a case study of Open Data Pakistan

**Nosheen Fatima Warraich,** and **Tayyba Rasool**
Institute of Information Management,
University of the Punjab,
Lahore, Punjab, PAKISTAN
e-mail: nosheen.im@pu.edu.pk (corresponding author);
tayyaba.rasool@yahoo.com
ORCID ID: N.F.Warraich: 0000-0002-1901-9743
T.Rasool: 0000-0001-7271-4190

**ABSTRACT**

*Open data portals have been implemented in numerous countries to provide accessible, usable, and reproducible data. This study aims to assess the readability of datasets from Open Data Pakistan. Using the URL (https://opendata.com.pk/) in an online readability checking tool, six well-known readability formulas were employed, including Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, Smog Index, Coleman Liau Index, and Automated Readability Index. The average score of each formula was calculated to analyze the standard readability of the open data portal. Additionally, the website was explored to evaluate the overall status of open datasets, showcases, and features like "Connect" on Open Data Pakistan. The findings indicated that the Open Data Pakistan website has a reading difficulty, scoring an average of 47.4 out of 100, making it challenging for the general public to read. However, the overall analysis suggested that the readability status of open data in Pakistan is satisfactory. These results offer insights into improving the readability of open data for enhanced citizen participation and informed decision-making. Moreover, the study has implications for policymakers and stakeholders regarding the readability of Open Data Pakistan, emphasizing the importance of considering readability in open data publication policies and guidelines, with a recommendation to decrease the percentage of open data portals categorized as hard to read.*

**Keywords:** Open Data Pakistan; Readability; Open data portal; Open government data; Digital transparency

## INTRODUCTION

In the age of data-driven decision-making and digital transparency, open data initiatives have become a crucial method to democratize information and encourage increased citizen engagement. Globally, governments, organizations, and institutions are adopting the idea of open data, providing non-sensitive data freely to the public through dedicated platforms called open data portals (Davies and Bawa 2012). These portals serve as gateways, providing citizens, researchers, policymakers, and businesses access to extensive digital repositories of information that can drive innovation, enhance public services, and promote the development of evidence-based policies (Kitchin 2014). It has been asserted that open government data (OGD) plays a crucial role in fostering social control over the implementation of public policies, the utilization of public resources, government

transparency, efforts to combat corruption, legal actions, and the development of cutting-edge data-driven applications for social good (Kozievitch et al. 2022). Given the increasing significance of open data portals as key resources for harnessing actionable insights, it becomes imperative to assess their effectiveness in adhering to fundamental information principles such as availability, accessibility, and readability (Zuiderwijk et al. 2012).

The readability of a written piece influences how easy or challenging it is to read and comprehend. Data processing transforms data into information, and this information is only valuable when it can be effectively understood. The readability measure, a fundamental mathematical formula, plays a key role in predicting the reader's comprehension level of written content. Assessing readability provides an indication of how easily the content can be understood by the reader. A text document's readability score affects both the content's accessibility and reading speed. Poor-quality text in a document can result in longer comprehension times for readers (Akgül 2022). A previous research has explored the use of data visualizations and data schemas in facilitating better comprehension and utilization of open data (Kitchin 2014). Evaluating the readability of web pages is crucial to ensure that the content is accessible to a diverse audience (Ismail et al. 2019). It is particularly crucial to design government websites with usability, accessibility, and readability in mind, as these websites aim to inform potential users about e-services. Websites with poor design have a negative impact on usage and encourage less online interaction (Akgül 2019; Baker 2009; Clemmensen and Katre 2012).

Open data has become commonplace in contemporary times. Numerous countries are increasingly introducing OGD portals, providing accessible data for individuals to use for their respective purposes. OGD fosters collaboration among government departments, civil society organizations, academia, and the corporate sector, facilitating the co-creation of solutions to social challenges through the utilization of shared knowledge and skills. OGD embodies the concept of making publicly created data available in a machine-readable and easily accessible format. It entails making available datasets, publications, and other documents and materials gathered and maintained by government entities, with no constraints on reuse. Transparency, accountability, and citizen involvement are the fundamental goals of OGD. While open data holds the potential for economic value in the millions and billions, not all open data is effectively utilized or repurposed (Nikiforova 2021).

Given that OGD enables citizens to monitor the effectiveness and leadership of the government, it not only facilitates decision-making grounded in facts but also directly influences individuals' perceptions of the government and their trust in it. Governments are considered to be the primary benefactors of "GovTech," wherein government transparency and OGD policies are being created to facilitate citizens to access information and engage with the government. Civic Tech, on the other hand, consists of several initiatives that use OGD to serve the general welfare (Yoshida and Thammetar 2021). This necessitates the openness of "correct" or "accurate" data, i.e., information that will be valuable to consumers as well as to governments that will enable data opening and take advantage of it.

Websites are assessed for their readability, performance quality, and usability. Trust is crucial for accessing e-government services, as it facilitates the effective utilization of websites for e-government purposes (Huang et al. 2009). However, e-government websites often face challenges related to usability, accessibility, and readability (Ho 2002; Youngblood and Mackiewicz 2012). Ensuring the readability criteria of these websites is of

utmost importance. If the contents offered are of high quality and easy for the great majority of people to grasp, the Internet might remove obstacles to public access to quality information and, as a result, reduce disinformation. Websites operated by the government or other organizations should be regarded as trusted sources for the general public. The readability and quality scores of these websites are crucial, as high-quality and easily understandable content can eliminate barriers to public access to health information, potentially reducing misinformation (Fogel et al. 2001).

Pakistan increasingly recognizes the potential of open data as a catalyst for enhancing accountability, promoting transparency, and utilizing data to advance societal progress (Saxena and Muhammad 2018). The Pakistan Citizen's Portal is one of the open data initiatives that aim to solve issues of corruption, lack of transparency, and low public trust in government institutions (Government of Pakistan 2019). The multifaceted effects of open data, such as their role in promoting socioeconomic development (Ahmed, Mahmood and Hu 2018) and enabling data-driven governance (Iqbal and Shaikh 2021), highlighted the importance of studying open data in Pakistan.

Additionally, readily accessible data can play a pivotal role in advancing vital sectors such as healthcare and education. The availability of data holds the promise of fostering innovation, improving patient care, and shaping informed policies. An illustrative example is the utilization of national health data, which played a central role in numerous countries' efforts against COVID-19, aiding governments in easing restrictions and steering toward a return to normalcy. Research within the open data initiative holds significance in shaping government policies and programs, promoting more open and accountable governance, and ultimately advancing societal well-being. Notably, the literature lacks sufficient coverage of the evaluation of open data in Pakistan. This study aims to fill this gap by evaluating the readability of datasets from "Open Data Pakistan." It also measures status of datasets, showcases, the facility for comments, complaints, feedback, and suggestions in an open data portal. This study poses the following research question: *What is the readability status of datasets from Open Data Pakistan?*

## LITERATURE REVIEW

The dissemination of information services to individuals is facilitated with greater ease, speed, and efficacy through web-based applications and the Internet, contributing to enhanced organizational effectiveness (Verkijika and De Wet 2018). Websites have assumed a novel and integral role in facilitating communication between governments and their constituents, serving as a platform and medium for information sharing, improved accessibility, service delivery, and transformative interactions with the public, other government branches, corporations, and various stakeholders (Jun, Wang and Wang 2014; Pérez-López, Prior, and Zafra-Gómez 2015).

Data catalogues are essential in today's data-driven businesses because they make it easier to find, comprehend, and use a variety of data resources. However, it can be difficult to guarantee their dependability and quality, particularly in open contexts with lots of data. In addressing the need for effective and reliable quality assessment processes, Martinez-Gil (2023) proposed a methodology to automatically analyze the quality of open data catalogues. The study explored various approaches for evaluating compatibility and similarity between these catalogues, incorporating non-core quality dimensions such as provenance, readability, and licensing. Additionally, the core quality dimensions

considered encompassed accuracy, completeness, consistency, scalability, and timeliness. The overarching goal was to empower data-driven organizations to make informed decisions by leveraging dependable and thoughtfully selected data assets.

Public access to official data is facilitated by OGD, which is based on the principles of transparency, accountability, honesty, and integrity. Utilizing the information systems success model as a theoretical framework, Almuqrin et al. (2023) investigated the impact of individuals' trust in OGD on their perceptions of data, system, and service quality. The findings affirmed the influence of data, system, and service quality on citizens' perceptions of OGD credibility. Notably, the primary driver of residents' trust in OGD was identified as service quality, which, in turn, affected data and system quality. This underscores the significance of open government platforms in delivering public services, providing accurate and comprehensive data, incorporating feedback mechanisms, and employing data visualization for users.

The majority of European nations adhere to the Public Sector Initiative (PSI), which provides guidelines for open data and openness. Notably, Schauppenlehner and Muhar (2018) emphasize that only one country, i.e. Austria, actively addresses data use and provides strategies to ensure widespread and free accessibility of data to public bodies in society. Emphasizing the critical importance of metadata completeness and quality, Schauppenlehner and Muhar (2018) explained how the presence of metadata services alone does not ensure swift access to data and information. Open data policies, usually encompassing guidelines for making data open, accessible, and transparent, often fall short. This current scenario tends to restrict access to open data primarily for experts rather than facilitating broad public access. In terms of the self-declared objectives aimed at contributing to societal processes, both platforms were found to perform inadequately.

In their study, Fernández et al. (2021) examined the COVID-19 open data provided by the Spanish regions and the Ministry of Health. The findings revealed that both the Ministry of Health and 15 regions in Spain shared open data related to COVID-19. While the information presented varied across autonomous communities on the open data portals, the content encompassed PCR-verified data. Despite the abundance of data available in reusable formats, users were required to consult multiple sources to gain a comprehensive understanding of the national epidemiological situation.

Transparency or openness, often measured by the extent of information disclosure on public websites, is crucial for maintaining transparency across all levels of public values. Governments contribute to this by posting governance-related materials online, including financial statements and relevant laws. However, e-government websites, with only 1% of documents or publications available, performed poorly in terms of the volume of material provided. On the positive side, websites received the highest rating, achieving 97% of the possible points, for their ability to offer a diverse range of publications and documents (Akgül 2022).

## Readability of Open Data

Readability metrics have been applied to various domains, including academia, patents, newspapers, government websites, and medicine. Assessing readability becomes particularly crucial when considering the comprehension level of a text document or web page for individuals with lower literacy skills. In such cases, the complexity of texts, characterized by polysyllabic words and lengthy, intricate sentences, can pose challenges for less experienced writers (Akgül 2021). Guidelines for web content readability are

provided by W3.0rg, aiming to make text material accessible to users and assistive technology while ensuring the availability of necessary information for interpretation.

In their study, Risoldi Cochrane, Gregory, and Wilson (2012) conducted a comparison of the readability of healthcare data for consumers on websites supported by the U.S. government and those supported by private businesses. The assessment utilized three validated metrics—SMOG Formula, Flesch Reading Ease Score (FRES), and Flesch-Kincaid Reading Level (FKRL) —to evaluate the readability of the web pages. Mann-Whitney U test was applied to compare the average readability of websites supported by the government versus those supported by businesses. According to the FKRL and FRES, commercially financed websites were much more challenging to read. According to the SMOG Formula, there was no significant change. Consumer-oriented health information on the Internet generally exhibited poor readability. Additionally, an evaluation of the readability of Indian open government data showed that more than 43.28 percent of the data had language that was challenging to comprehend (Ojha, Ismail and Kuppusamy 2018). Turkish e-government websites demonstrated a notably low level of readability, with state and local governments' websites receiving a FRES score indicating "difficult to read" (Akgül 2019).

In Akgül's (2022) examination of the usability, readability, and public values of Turkish national-level open government websites, the findings revealed an average Gunning Fog Index (GFI) score of 12.79. This score indicated reading levels suitable for the average college graduate. Similarly, Yeung et al. (2022) assessed the readability and quality of online content related to COVID-19 immunization on official/governmental websites. The typical FRES and FKRL of frequently asked questions about vaccinations on these websites were 40.9 and 12.1, respectively. The findings indicated that the open data portals for COVID-19 vaccination were not easily readable. In a recent study, Serry et al. (2023) evaluated the readability of web pages from two Victorian government websites designed for the general public, disseminating crucial health information during the 2020 COVID-19 pandemic. The observed text difficulty exceeded the typical levels for health promotion materials targeted at senior primary school levels. This finding was consistent across both public and professional target audiences, indicating that effective engagement with the posted text required reading at the senior secondary level. Ismail et al. (2019) conducted an analysis of the site rankings, readability, and accessibility for the top 20 government websites in India (N = 20). Six reliable strategies were employed to assess the readability of the websites' content. These websites' readability scores were found to be within acceptable bounds. However, considering that the results are derived from the United States grading scale, the development of a national grading system for readability is essential.

## METHOD

This study focused on evaluating the datasets available on "Open Data Pakistan." The importance of easily accessible government data cannot be overstated in today's context. Government websites, particularly those hosting open data sets, play a crucial role in providing the public with essential information about government operations. Open Data Pakistan (https://opendata.com.pk) is a prominent advocate for OGD and transparency, with datasets covering various aspects of the nation's economy, demography, and governance. Covering diverse topics such as healthcare, education, and agriculture, the platform offers a wide array of data to cater to a broad range of interests. This reflects the

government's commitment to openness and accountability. The readability of datasets was assessed using measures such as FKRL and FRES.

This readability metric, which is most frequently used, assigns an understanding level to the delivered English text on a scale from 0 to 100 points. The higher the score, the easier it is to read and comprehend the information or data; conversely, the lower the score, the more challenging it is to understand the content. Higher grades indicate simpler reading material, with a typical 11-year-old easily understanding scores ranging from 90 to 100. Scores ranging from 60 to 70 are regarded as typical readability, easily comprehended by children between the ages of 13 and 15. Lower grades imply complex readings (Flesch 1948). Additionally, the Smog Index, Coleman Liau Index, and Automated Readability Index were also used to evaluate the readability of the open data portal of Pakistan. Several previous studies also applied the Flesch-Kincaid Grade Level and Flesch Reading Ease, Gunning Fog Score, Smog Index, Coleman Liau Index, and Automated Readability Index to check the readability of open data portals (Akgül 2022; Ismail et al. 2019; Ojha, Ismail and Kuppusamy 2018; Risoldi Cochrane, Gregory, and Wilson 2012).

The readability findings for the open data portals are produced by the online readability checking tool using the URL of the open data portal as input in May 2023. The readability findings for the open data portals were generated using an online readability checking tool with the URL of the open data portal as input in May 2023. The tool output values for six well-known readability formulas, including Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, Smog Index, Coleman Liau Index, and Automated Readability Index, based on the Open Data Pakistan URL (https://opendata.com.pk/). The average scores of each formula were then calculated to assess the standard readability of the open data portal.

Open Data Pakistan's virtual tour was an engaging experience. Researchers were astounded by the website's enormous collection of datasets as they browsed around, each of which held the key to revealing insights into different facets of the nation's economy, demography, and governance. The necessity of evaluating the platform's content readability became more and clearer as researchers dug further into it. In this sense, readability refers to the ability of the data and information to be understood and accessed by a wide range of users. It is essential to make sure that the data is accessible to the public, researchers, and policymakers, and that they can use it. Further, the URL of Open Data Pakistan was used as input to generate readability findings in the online readability checking tool.

Exploring Open Data Pakistan's virtual tour was a captivating experience for the authors, who were impressed by the extensive collection of datasets offering insights into various aspects of the nation's economy, demography, and governance. As the authors delved deeper, the importance of evaluating the readability of the platform's content became increasingly evident. In this context, readability refers to the data and information's ability to be comprehended and accessed by a diverse audience, including the public, researchers, and policymakers. Ensuring that the data is easily accessible and usable is crucial. The online readability checking tool utilized the URL of Open Data Pakistan to generate readability findings. The authors assessed the quantity of datasets, showcases (including applications, websites, and visualizations), and the overall status of open data by visiting the website in May 2023. Additionally, each showcase was thoroughly examined to address the research question posed in the study.

## RESULTS

### The Landscape of Open Data in Pakistan

The study identified a total of 837 datasets, comprising 818 secondary research datasets and 19 primary research datasets. These datasets cover 14 categories, including among others Public Safety, Economy & Finance, Health, and Education, and originate from various locations in Pakistan, such as Khyber Pakhtunkhwa (KPK), Sindh, Baluchistan, Islamabad, Lahore, Gilgit Baltistan, and Azad Jammu and Kashmir. The datasets are available in various formats, such as CSV (439), XLSX (198), URL (19), XLS (14), DOCX (5), SAV (2), DO (1), DTA (1), ZIP (50), PDF (173), HTML (1), PNG (1), PPT (1), and RAR (1).

Moreover, the study identified 13 showcases covering various topics such as sports, health, gender, attacks, environment, climate, crime reporting, consumer, residential electricity consumption dataset, brain drain, COVID-19, and suicide bombings. These findings highlight the richness of Open Data Pakistan with a diverse array of datasets across various fields. Users can explore the full datasets page by either launching the website/preview or directly clicking on the download button. The platform offers features to view, download, share, add to favorites, and nominate for Viz of the day (visualization each day) for a showcase. While users can see the number of views for a showcase, the number of downloads and shares remains undisclosed. Additionally, comments can be viewed if a dataset has received user comments. These datasets are freely accessible to the public, facilitating individuals, researchers, and organizations in accessing and analyzing the data. The provision of open data contributes to transparency, accountability, and evidence-based decision-making in Pakistan. Moreover, the open data provides the facility of "connect" for feedback, complaints, and suggestions regarding open datasets. However, this feedback or complaints are not visible to other viewers.

### Readability Metrics and Readability Assessment of Open Data Pakistan

An assessment of readability gauges how easily content can be comprehended by the reader. The readability score of a text document influences the reading speed and the ease with which readers grasp its content. When the text in a document is inadequate, it hinders reader comprehension. The readability score was determined using the six formulas from the online readability checker tool, including the Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, Smog Index, Coleman Liau Index, and Automated Readability Index. This was done to analyze the readability of Open Data Pakistan.

### (a) Flesch–Kincaid Reading Ease (FKRE)

This readability metric, which is most frequently used, assigns an understanding level to the delivered English text on a point scale ranging from 0 to 100 (Table 1). The higher the score, the easier it is to read and comprehend the information, whereas a lower score indicates that the content is more challenging to understand. Higher grades indicate simpler reading material, with an average 11-year-old easily understanding scores between 90 and 100. Scores between 60 and 70 are regarded as typical readability, easily understood by children between the ages of 13 and 15.

The readability score of FKRE is 47.4, which indicates that Open Data Pakistan's content is hard to read for the general citizens. Existing literature aligns with this finding, suggesting that the FKRE score surpassed the average cutoff value (Akgül 2019, 2022; Ismail et al. 2019; Serry et al. 2022).

Table 1: Flesch–Kincaid Reading Ease (FKRE): Text Readability Point Scale

| Readability Score | Understandability level |
|---|---|
| 90–100 | Very Easy |
| 80–89 | Easy |
| 70–79 | Fairly Easy |
| 60–69 | Standard |
| 50–59 | Fairly Difficult |
| 30–49 | Difficult |
| 0–29 | Very Confusing |

## (b) Flesch–Kincaid Grade Level (FKGL)

The essential metrics used by the Flesch-Kincaid Grade Level (FKGL) and FRES are the same, however, a different weighting factor is used. An average eighth-grade student is thought to understand the topic if they receive a score of 8.2 (Kincaid et al. 1975). This method serves as a benchmark test for the US government's Defence department. This formula considers any grade value above 12 as equivalent to a grade value of 12. A score of 5.0 likely implies grade school level and a score of 7.4 indicates that a typical 7th-grade kid can understand the content (Ismail et al. 2019). The FKGL's score of Open Data Pakistan is 7.8 indicating that the data readability is good. It provides evidence that readability is easily comprehended by the general public. The results are consistent with the findings of Ismail et al. (2019) and the study conducted by Risoldi Cochrane, Gregory, and Wilson (2012). Nevertheless, other studies have reported that the FKGL score may be challenging for the general public in terms of readability (Akgül 2019, 2022; Serry et al. 2022).

## (c) Gunning Fog Index (GFI)

The Gunning Fog Index (GFI) determines how many years of education are required to understand an English text after just one reading. The text readability statistic is based on the number of difficult words and the length of the sentences. A typical index would have six for the Bible, 10 for Time Magazine, fourteen for The Times Newspaper, and above than fifteen academic publications. Anything above 12 indicates texts that are too challenging for the majority of readers to understand, with a score of 7-8 being optimal (Gunning 1952).

Similar to the Flesch scale, the Fog Index measures words with three or more syllables based on their name. Anything above 12 on this index is too difficult to read, and a score of 7 or 8 is appropriate. That is, in general, a score of 5 is readable, 10 is hard, 15 is difficult and 20 is very difficult to grasp the text. It estimates the number of years of formal education required for first-time reading comprehension (Ismail et al. 2019).

The findings of the average GFI readability assessment revealed that the average GFI score was 6.2. This suggested reading levels that a typical sixth grade could understand. This GFI score indicated that the readability of Open Data Pakistan is easy to understand even for citizens having a sixth-grade education level based on word measures. The findings are also supported by Ismail et al. (2019). However, the others, (Akgül 2019, 2021, 2022) found that the GFI score was insignificant indicating hard to read for the general public.

## (d) SMOG Index

The SMOG, or "Simple Measure of Gobbledygook," is an acronym for a readability test that assesses word length to gauge the level of understanding required for readers to comprehend written content. The recommended score is 7, according to the WCAG (Web

Content Accessibility Guidelines). The difficulty of reading grows as the number rises (Henry 2018).

The SMOG Index formula is thought to be suitable for secondary-age readers, or readers from the fourth grade to college level. The outcome, which is based on grade levels used in US schools, suggests that the average student who can read the material fits inside that scale. For instance, the typical 7th-grade student can understand the material with a score of 7.4 (Ismail et al. 2019). The score of the SMOG index for the Open Data Pakistan is 6.2. It indicated that the written text is easy to comprehend by secondary-age readers (13 and 15 age group). The findings are also supported by Akgül et al. (2019) and Ismail et al. (2019). However, Serry et al. (2022) findings did not support a minimum of 8.2, showing that the written text was hard to read for secondary-age readers. Further, Risoldi Cochrane, Gregory, and Wilson's (2012) study showed SMOG score was insignificant for the readers.

### (e) Coleman Liau Index (ClI)

The Coleman Liau Index (CLI) relies on letters instead of syllables per word and sentence length. It utilizes a US grade-based system to determine the readability of the text. According to the character-based formula proposed by Coleman and Liau (1975), computerized evaluations of character understanding are easier and more accurate than counting syllables and sentence length.

The CLI score for the Open Data Pakistan is 14.8. This score indicated that the text was too hard for the majority of the readers. The findings are also supported by the literature (Ismail et al. 2019).

### (f) Automated Readability Index (ARI)

The Automated Readability Index (ARI) is calculated based on ratios that describe the difficulty of words and sentences. In addition to basing its output on the US grading level system, ARI provides a number that roughly indicates the age required to understand or comprehend the content (Kincaid et al. 1975). It indicates that an ARI output of 10 corresponds to a high school student, typically aged 15 to 16 years old, while a score of 3 suggests that third graders, who are 8 to 9 years old, should be able to understand the material.

The ARI output score for Open Data Pakistan is 5.3. It indicated that students in grade 5 and above, typically aged 10-11 years, can comprehend the text. This also suggests that individuals with a 5th-grade education can read the datasets of Open Data Pakistan. The findings are also supported by Ismail et al. (2019).

## DISCUSSION

Open Data Pakistan (https://opendata.com.pk/) promotes data accessibility, usability, and sharing for businesses, governments, and individuals, fostering improvements in social, economic, and environmental aspects. This study explores the readability status of open data in Pakistan, utilizing FKGL and FRES to assess the readability. These readability metrics are widely recognized and employed for evaluating open data portals (Akgül 2022; Ismail et al. 2019; Ojha, Ismail and Kuppusamy 2018; Risoldi Cochrane, Gregory, and Wilson 2012).

The readability evaluation summary for Open Data Pakistan is presented in Table 2. The analysis indicates that the FKRE score is 47.4, suggesting that the content may pose a challenge for the general public. The readability tool generated an overall score for the websites' readability, and Open Data Pakistan received an overall average readability score of 47.4, with 100 being the highest readability score. The low score from the readability tool suggests that the content on ODP is challenging for the average citizen to comprehend. According to readability recommendations for web content (Flesch 1948), a score of at least 60 on a scale of 0 to 100 is considered acceptable. A text with a reading score of 60 to 70 is comparable to a grade level of 8 to 9, meaning it should be understandable for individuals aged 13 to 15. Additionally, the FKGL score for Pakistan's open data is 7.8, indicating good readability and ease of understanding for a general audience.

Furthermore, the average GFI readability evaluation results indicate that Open Data Pakistan has an average GFI score of 6.2. This score, based on word metrics, suggests that the readability of Open Data Pakistan is easily comprehensible even for individuals with a sixth-grade education level. Open Data Pakistan's SMOG index score is 6.2, indicating that secondary readers (aged 13 to 15) can easily understand the written text. However, the CLI score of 14.8 suggests that the text may be too challenging for the majority of readers. The ARI output score of 5.3 indicates that students in grades 5 and above (approximately 10 to 11 years old) can comprehend the text.

Table 2: Summary of the Readability Findings for Open Data Pakistan

| Readability measures | Readability score | Level | Age | Grade Level |
|---|---|---|---|---|
| Flesch–Kincaid Reading Ease (FKRE) | 47.4 | Very Confusing | 17 and above | College level |
| Flesch–Kincaid Grade Level (FKGL) | 7.8 | Good | 12 and above | 7th Grade |
| Gunning Fog Index (GFI) | 6.2 | Easy to understandable | 11 and above | 6th Grade |
| SMOG Index | 6.2. | Easy to comprehend | 13 and 15 | 9th Grade |
| Coleman Liau Index (ClI) | 14.8 | Too hard to read | 17 and above | 11th Grade |
| Automated Readability Index (ARI) | 5.3. | Understandable | 10 and 11 | 5th Grade and above |

These findings imply that Open Data Pakistan may require a higher level of reading comprehension, catering to readers with advanced degrees or specialized subject expertise. Simplifying the writing or providing more explanations could be beneficial for readers less familiar with the subject. Open data plays a crucial role in improving the understanding of government operations and decision-making processes, allowing people, researchers, and businesses to access and analyze government data. Governments can streamline operations and enhance service delivery through data-driven policy formation and evidence-based decision-making. Additionally, open data fosters the development of innovative services and products that benefit society as a whole.

## CONCLUSIONS

This study assessed the readability, showcasing, and status of datasets on the Open Data Pakistan portal, including its capacity for comments and feedback. As of May 2023, the portal hosts 837 datasets across 14 categories, available in various formats. Users can access, view, share, and download datasets. The "Connect" tab allows users to provide feedback and suggestions. The readability check revealed an overall score of 47.4%, indicating that the content is challenging to read. To improve accessibility, incorporating easy-to-read phrases in web material is recommended, aiming for readability suitable for 13 to 14-year-olds.

This study is confined to evaluating the readability of Open Data Pakistan and does not encompass content or usability analysis. The FKGL and FRES were employed for readability assessment, a widely recognized scale for open government portals. However, the reliance on an online tool for Open Data Pakistan's readability evaluation may be considered a limitation in this study.

The study may shed light on the efficacy of these initiatives in engaging the public by evaluating the readability of open data in Pakistan. It can contribute insights on enhancing the accessibility of open data to foster active participation and empower citizens for informed decision-making. Additionally, the study holds implications for stakeholders and policymakers, underscoring the importance of considering readability in data publication policies. It opens avenues for discussions on readability, accessibility, and usability of open data portals in Pakistan, potentially leading to the establishment of standards, guidelines, or best practices.

## ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or non-for-profit sectors.

## AUTHORS DECLARATION

The authors declare no conflicts of interest regarding the publication of this paper.

## REFERENCES

Ahmed, M., Mahmood, A. N., and Hu, J. 2018. Role of open data in social development: a case study of Pakistan. In *Proceedings of the 19th Annual International Conference on Digital Government Research* (pp. 26-34).

Akgül, Y. 2019. The accessibility, usability, quality, and readability of Turkish state and local government websites an exploratory study. *International Journal of Electronic Government Research* (IJEGR), Vol. 15, no. 1: 62-81.

Akgül, Y. 2021. Accessibility, usability, quality performance, and readability evaluation of university websites of Turkey: a comparative study of state and private universities. *Universal Access in the Information Society*, Vol. 20, no. 1: 157-170.

Akgül, Y. 2022. Evaluating the performance of websites from a public value, usability, and readability perspective: a review of Turkish national government websites. *Universal*

*Access in the Information Society.* Available at: https://doi.org/10.1007/s10209-022-00909-4.

Almuqrin, A., Mutambik, I., Alomran, A., Gauthier, J., and Abusharhah, M. 2022. Factors influencing public trust in open government data. S*ustainability*, Vol. 14, no. 15: 9765.

Baker, D. l. 2009. Advancing e-government performance in the United States through enhanced usability benchmarks. *Government Information Quarterly*, Vol. 26, no. 1: 82–88. Available at: doi:10.1016/j.giq.2008.01.004.

Clemmensen, T., and Katre, D. 2012. Adapting e-gov usability: evaluation to cultural contexts. In E. Buie, & D. Murray (Eds.), *Usability in Government Systems: User Experience Design for Citizens and Public Servants,* pp. 331-344. Elsevier. Available at: https://doi.org/10.1016/B978-0-12-391063-9.00053-5.

Coleman, M., and Liau, T. L. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, Vol. 60, no. 2: 283.

Davies, T. G., and Bawa, Z. A. 2012. The promises and perils of open government data (OGD). *The Journal of Community Informatics*, Vol. 8, no. 2.

Fernández, A. M., Latorre, M. A., Sánchez-Eznarriaga, B. C., and García, L. F. 2021. Availability of open data related to COVID-19 in the web portals of the Spanish regions and the Ministry of Health. *Revista Espanola de Salud Publica*, Vol. 95, e202112190.

Flesch, R. 1948. A new readability yardstick. *Journal of Applied Psychology*, Vol. 32, no. 3: 221.

Fogel, J., Albert, S. M., Schnabel, F., Ditkoff, B. A., and Neugut, A. I. 2001. Quality of health information on the Internet. *JAMA*, Vol. 286, no. 17: 2092-2095.

Government of Pakistan. 2019. Pakistan Citizen's Portal: annual performance report 2018-2019. Available at: https://www.citizensportal.pk/downloads/annual-performance-reports/AnnualPerformanceReport-2018-2019.pdf.

Gunning, R. 1952. *The technique of clear writing*. New York: McGraw-Hill.

Henry, S.L. 2018. Referencing and linking to WAI guidelines and technical documents. Available at: https://www.w3.org/WAI/standa rds-guidelines/linking/.

Ho, A. T. K. 2002. Reinventing local governments and the e-government initiative. *Public Administration Review*, Vol. 62, no. 4: 434-444.

Huang, Z., Brooks, L., and Chen, S. 2009. The assessment of credibility of e-government: users' perspective. In: Salvendy, G., Smith, M.J. (eds) *Human Interface and the Management of Information. Information and Interaction.* Human Interface 2009. Lecture Notes in Computer Science, vol 5618. Springer, Berlin, Heidelberg. Available at: https://doi.org/10.1007/978-3-642-02559-4_4.

Iqbal, M. S., and Shaikh, K. 2021. Data-driven governance in Pakistan: a comparative analysis of the Ehsaas program and the Citizen's Portal. In *Proceedings of the 22nd Annual International Conference on Digital Government Research,* pp. 71-79.

Ismail, A., Kuppusamy, K. S., Kumar, A., and Ojha, P. K. 2019. Connect the dots: accessibility, readability and site ranking–an investigation regarding top-ranked websites of the Government of India. *Journal of King Saud University - Computer and Information Sciences*, Vol. 31, no. 4: 528-540.

Jun, K. N., Wang, F., and Wang, D. 2014. E-government use and perceived government transparency and service capacity: evidence from a Chinese local government. *Public Performance & Management Review*, Vol. 38, no. 1: 125-151.

Kincaid, J. P., Fishburne Jr., R. P., Rogers, R. L., and Chissom, B. S. 1975. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. *Institute for Simulation and Training,* 56. Available at: https://stars.library.ucf.edu/istlibrary/56/.

Kitchin, R. 2014. The data revolution: Big data, open data, data infrastructures, and their consequences. Thousand Oaks, CA: Sage Publications Ltd.

Kozievitch, N. P., Fonseca, K. V., Rosa, M. D. O., Berardi, R., Gutierrez, M. B., Belizario, M., Luz, M.G. and Liu, E. 2022. Assessment of Open Data Portals: a Brazilian case study. *IEEE International Smart Cities Conference (ISC2)*, Pafos, Cyprus, 2022, pp. 1-7, Available at: doi: 10.1109/ISC255366.2022.9921914.

Martinez-Gil, J. 2023. Framework to automatically determine the quality of open data catalogs. *arXiv preprint,* Available at: arXiv:2307.15464.

Nikiforova, A. 2021. Smarter open government data for Society 5.0: Are your open data smart enough? *Sensors*, Vol. 21, no. 15: 5204.

Ojha, P. K., Ismail, A., and Kuppusamy, K. S. 2018. Readability assessment-cum-evaluation of government department websites of Rajasthan. In *Proceedings of First International Conference on Smart System, Innovations and Computing*: SSIC 2017, Jaipur, India, pp. 235-244. Springer Singapore.

Pérez-López, G., Prior, D., and Zafra-Gómez, J. L. 2015. Rethinking new public management delivery forms and efficiency: Long-term effects in Spanish local government. *Journal of Public Administration Research and Theory*, Vol. 25, no. 4: 1157-1183.

Risoldi Cochrane, Z., Gregory, P., and Wilson, A. 2012. Readability of consumer health information on the internet: A comparison of US government–funded and commercially funded websites. *Journal of Health Communication*, Vol. 17, no. 9: 1003-1010.

Saxena, S. and Muhammad, I. 2018. The impact of open government data on accountability and transparency. *Journal of Economic and Administrative Sciences*, Vol. 34, no. 3: 204-216.

Schauppenlehner, T., and Muhar, A. 2018. Theoretical availability versus practical accessibility: The critical role of metadata management in open data portals. *Sustainability*, Vol. 10, no. 2: 545.

Serry, T., Stebbins, T., Martchenko, A., Araujo, N., and McCarthy, B. 2023. Improving access to COVID-19 information by ensuring the readability of government websites. *Health Promotion Journal of Australia*, Vol. 34, no. 2: 595-602.

Verkijika, S. F., and De Wet, L. 2018. Quality assessment of e-government websites in Sub Saharan Africa: A public values perspective. T*he Electronic Journal of Information Systems in Developing Countrie*s, Vol. 84, no. 2: e12015.

Yeung, A. W. K., Wochele-Thoma, T., Eibensteiner, F., Klager, E., Hribersek, M., Parvanov, E. D.,Hrg, D., Völkl-Kernstock, S., Kletecka-Pulker, M., Schaden, E., Willschke, H. and Atanasov, A. G. 2022. Official websites providing information on COVID-19 vaccination: readability and content analysis. *JMIR Public Health and Surveillance*, Vol. 8, no. 3: e34003. doi: 10.2196/34003.

Yoshida, M., and Thammetar, T. 2021. Education between govtech and civic tech. *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 16, no. 4: 52-68.

Youngblood, N. E., and Mackiewicz, J. 2012. A usability analysis of municipal government website home pages in Alabama. *Government Information Quarterly*, Vol. 29, no. 4: 582-588.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., and Alibaks, R. S. 2012. Socio-technical Impediments of Open Data. *Electronic Journal of e-Government,* Vol. 10, no.2: 156-172.