Unveiling Open Data Pakistan: Assessing the Availability, Accessibility, and Readability of Open Data Portal

Nosheen Fatima Warraich, and **Tayyba Rasool**
Institute of Information Management,
University of the Punjab,
Lahore, Punjab, PAKISTAN
e-mail: nosheen.im@pu.edu.pk; tayyaba.rasool@yahoo.com

ABSTRACT
Open data portals are being introduced by many countries, offering data that anybody may access and use and repurpose. This study aims to evaluate the readability of datasets from "Open Data Pakistan" along with their availability and accessibility. The URL of Open Data Pakistan (https://opendata.com.pk/) was used as an input in an online readability checking tool (https://www.webfx.com/tools/read-able/) which produced the values of six well-known readability formulas as output 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 determined to analyze the standard readability of the open data portal. Further, the website was explored to evaluate the overall status of open data sets, showcases, and other features i.e. "Connect" of Open Data Pakistan. The findings indicated that the Open Data Pakistan website has a reading difficulty of 47.4 out of 100 on average. It indicated that it is hard for the general public to read. Further, the overall analysis of the portal directed that citizens have satisfactory access to and availability of open data. These findings provide insights into how open data can be made more accessible to encourage active participation and empower citizens to make informed decisions. Further, this study has implications for policymakers and stakeholders about the readability, availability, and accessibility challenges in open data. It also highlights the significance to consider readability and availability of open data publication policies and guidelines. It is recommended that the percentage of open data portals in the hard-to-read category be decreased, hence improving the information resources' accessibility.

Keywords: Open Data Pakistan, Availability, Accessibility, Readability, Open Data Portal

INTRODUCTION

In the era of data-driven decision-making and digital transparency, open data initiatives have emerged as a pivotal means to democratize information and foster greater citizen engagement. Governments, organizations, and institutions worldwide are embracing the

concept of open data, making non-sensitive data freely available to the public through dedicated platforms known as open data portals (Davies and Bawa 2012). These portals serve as gateways, providing citizens, researchers, policymakers, and businesses access to vast repositories of information that can drive innovation, enhance public services, and promote evidence-based policy development (Kitchin 2014). It has been asserted that open government data is crucial for fostering social control over the implementation of public policies, the use of general resources, government translucency, measures to combat corruption, legal action, and the development of cutting-edge data-driven apps for social good (Kozievitch et al. 2022). With the growing importance of open data portals as key resources for harnessing actionable insights, it becomes crucial to assess their effectiveness in adhering to fundamental principles such as availability, accessibility, and readability (Zuiderwijk et al. 2012).

Availability, the first critical aspect of public data portals, revolves around exactly how much these platforms offer comprehensive and up-to-date datasets. Institutions, organizations, and individuals can all find, share, combine, and reuse government information material thanks to the expanding availability and ongoing development of information technologies (Dawes 2010). According to Gurstein (2013), information should be made freely available in its whole, in inconvenient and flexible machine-readable formats, and for no more than the cost of its replication. Assessing the availability of data on open data portals is essential for understanding the breadth of information available to the public and identifying areas where improvements may be necessary (Zuiderwijk et al. 2015).

The second aspect of interest, accessibility, emphasizes the ease with which users can find, retrieve, and comprehend data presented on open data portals (Zuiderwijk et al. 2015). The goal of accessibility is to provide a way to get over barriers that prevent people from retrieving knowledge, provide photos with alternative text for those with eye impairments, give keyboard control as a choice to people with mobility difficulties, etc. (Lazar et al. 2010; Olalere and Lazar 2011; Reis et al. 2013). Studies have explored user experiences and challenges faced in navigating vast datasets, as well as language barriers and potential biases in data representation (Zuiderwijk et al. 2012).

The readability of a piece of writing determines how simple or complex it is to read and comprehend. Data processing produces information, and information is only useful when it can be understood. To make the contents accessible to a wide range of people, it is crucial to assess the readability of web pages (Ismail et al. 2019). A key mathematical formula for predicting the reader's level of comprehension of written content is the readability measure. An evaluation of readability describes how simple it is to give the reader content. A text document's readability score affects both the content's accessibility and reading speed. When a document's text is subpar, it takes the readers longer to understand it (Akgül 2022). Previous research has explored the use of data visualizations and data schemas in facilitating better comprehension and utilization of open data (Kitchin 2014). Designing websites with usability, accessibility, and readability in mind is especially important for government websites because they attempt 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 government data (OGD) encourages collaboration across government departments, civil society organizations, academia, and the corporate sector. It enables the co-creation of solutions to social difficulties through the use of communal knowledge and skills. OGD is the idea of making publicly available, publicly created data in a machine-readable and easily accessible manner. It entails making available datasets, papers, and other material gathered and maintained by government entities, with no constraints on reuse. Transparency, accountability, and citizen involvement are the fundamental goals of open government data. Even though open data has a potential economic value measured in the millions and billions, not all open data is recycled (Nikiforova 2021).

Open data is a common occurrence today. Open government data portals are being introduced by more and more countries, offering data that anybody may access and use for their purposes. Considering that open government data (OGD) enables citizens to track the effectiveness and leadership of the government, this not only supports decision-making based on facts but also directly affects how individuals perceive 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.

Researchers have looked into three more key issues that are mentioned in accessibility studies and that have an impact on users' access to websites that provide information. Websites are evaluated based on their readability, performance quality, and usability. Trust is essential for accessing e-government services since it enables websites to be utilized for e-government (Huang et al. 2009). However, usability, accessibility, and readability issues plague e-government websites frequently (Ho 2002; Youngblood and Mackiewicz 2012). These websites' readability, availability, and accessibility criteria are critical. 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 run by the government or other organizations should be trusted sources for the general public. These websites' readability and quality scores are very important because, if the presented materials are of high quality and simple to understand by the vast majority of people, the internet may remove obstacles to public access to health information and, as a result, eliminate misinformation (Fogel et al. 2001).

It was found that no study exists on the availability, accessibility, and readability evaluations of open data in Pakistan. Therefore, the evaluation of open data in Pakistan is not adequately covered in the literature. This study proposes to evaluate the readability of datasets of “Open Data Pakistan” along with the evaluation of the availability and accessibility of datasets,

showcases, and status of data. It also measures the facility for comments, complaints, feedback, and suggestions in an open data portal.
This study proposed the subsequent research questions:

RQ: What is the accessibility, availability, and readability status of datasets of Open Data Pakistan?

LITERATURE REVIEW

The transmission of information and services to people is easier, faster, and more effective thanks to Web-based apps and the Internet, which also improve organizational effectiveness (Verkijika De 2018). Websites now play a novel role in communication among governments and their constituents as a platform and medium for sharing information, better access, delivering service, and transformation interactions with the public, other branches of government, corporations, and other stakeholders (Jun et al. 2014; Pérez-López 2015).

Availability of Open Data

The Public Sector Initiative (PSI), which contains guidelines for open data availability, accessibility, and openness, is adhered to by the majority of European nations. The only country that addresses data use and offers tactics to guarantee that data is widely and freely available to public bodies in society (Schauppenlehner and Muhar 2018).

Fernández et al. (2021) examined the COVID-19 open data that had been made available by the Spanish regions, along with the Ministry of Health. It was discovered that the Ministry of Health, as well as 15 regions of Spain, disclosed open data relating to COVID-19. Although the information displayed varied from one autonomous community to another on the open data portals, the areas included information that was PCR verified. The data were found to be plentiful and in reusable formats, however, the user had to consult a variety of sources to learn about the epidemiological situation at a national level.

The completeness and quality of the metadata are extremely crucial, Schauppenlehner and Muhar (2018) explained how two significant European and Austrian metadata platforms demonstrated that quick access to data and information was not guaranteed by the mere existence of metadata services. Open data policies typically refer to guidelines for making open data available, accessible, and transparent. The existing situation results in restricted access to open data for experts instead of general public access. Concerning the self-declared goals of contributing to society processes, both platforms performed miserably.

The degree of information disclosure on public websites is referred to as transparency or openness. To maintain transparency and openness at all levels of public values, governments post governance-related material online, such as financial statements and pertinent laws. With only 1% of documents or publications available, e-government Websites did the lowest in terms of the volume of material provided there. Websites received the highest rating (97%

of the possible points) for their capacity to offer various publications and documents. It was determined that the development of open data portals depends on the availability of data via websites (Akgül 2022).

Accessibility of open data

The accessibility of Internet pages is one of the most crucial components of offering public Websites equal access for anybody who can see, understand, browse, and connect to the Internet, including those with impairments (Akgül 2019; Verkijika and De 2018). According to Wibowo et al. (2021), the most crucial category in the semantic web data quality models is accessibility. OGD portals' accessibility is crucial for users in all respects (Nikiforova and McBride 2021).

Akgül (2022) found that the accessibility of Turkish e-government Websites was far from acceptable. Further, Verkijika and De (2017) examined 217 e-government websites and discovered that none of them entirely met the accessibility requirements. The bulk of the websites were found to have numerous accessibility problems. Máchová et al. (2018) found accessibility as a barrier regarding national open data. This issue has decreased the open data usability among stakeholders. Likewise, Austrian and European open data portals were analyzed by Schauppenlehner and Muhar (2018), who discovered significant conceptual flaws and discrepancies that severely restrict practical accessibility. More, Abanikannda et al. (2017) investigated how agricultural science researchers in South-Western Nigeria used open data and discovered accessibility problems with it.

Gill and Corbett (2017) used heuristic evaluation rules to assess the British Columbia OGD portals' usability and accessibility from a design standpoint. They explained how there are major impediments to the accessibility and usefulness of open data when consumers have trouble locating and interacting with it on a portal. According to Olalere and Lazar (2011), many websites have accessibility issues that were not intended to be there in the first place but were later added to web pages. The majority of today's e-government websites regularly update and modify their web pages, which means that they are not static. When available, the site accessibility statements are particularly beneficial to users since they offer a road map for understanding the site's degree of compliance, its features, and the procedures used. The results confirmed that the homepage of the government open data portals did not violate any accessibility laws.

Additionally, there are no meaningful text equivalents and the storm preparation information on the front page is inaccessible to those with disabilities. Sheoran et al. (2023) showed how open data databases may offer high-quality information, for example, specific road systems and infrastructure that are available, which is helpful for accessibility analyses. Additionally, accessibility maps include information that is rather simple to grasp, which might aid the stakeholders in starting conversations about current issues and suggesting additional solutions.

Readability of open data

Readability metrics have been applied to a variety of fields, including academia, patents, newspapers, government websites, and medicine. When a low-literate person reads a text document or web page for the first time, they assist in determining the level of readability and ensure the level of understanding of the material. Texts with polysyllabic words and lengthy, complex sentences, for instance, penalize inexperienced writers (Akgül 2021). W3.0rg provided guidelines for the readability of web content. This standard aims to make text material readable by users and assistive technology while also ensuring that the information required to interpret it is readily available.

Risoldi et al. (2012) compared the readability of customers' healthcare data on websites supported by the U.S. government to that on websites supported by private businesses. Three verified metrics— SMOG Formula, Flesch Reading Ease, and Flesch-Kincaid Reading Level — were employed to assess the web pages' readability. 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 Flesch-Kincaid Reading Level and Flesch Reading Ease tests, 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 had poor overall readability. Further, An appraisal of the readability of Indian open government data revealed that over 43.28 percent of them have language that is difficult to understand (Ojha et al. 2018). Turkish e-government websites had a very poor level of readability, as evidenced by the fact that state and local governments' websites received a FRES score of "difficult to read'' (Akgül 2019).

Akgül (2022) looked at the usability, readability, and public values of Turkish national-level open government Websites. Findings exposed that 12. 79 was the average Gunning Fog Index (GFI) score. This suggested reading levels suitable for the average college graduate. Similarly, Yeung et al. (2022) evaluated and compared the content's readability and quality of online materials on COVID-19 immunization that were published on official/governmental websites. The websites' typical Flesch Reading Ease score and Flesch-Kincaid Grade Level that answer frequently asked questions about vaccinations were 40.9 and 12.1, respectively. It was determined that the OCVID-19 vaccination's open data portals were not easily readable. Ismail et al. (2019) provided an analysis of the site rankings, readability, and accessibility of the top 20 government websites in India (N = 20). Six reputable strategies are used to gauge how readable the website's material is. These websites' readability scores were found to be within acceptable bounds. However, because the results are based on the United States grading scale, a national grading system concerning readability must be developed. Serry et al. (2023) assessed the readability of web pages from two websites run by the Victorian government that were intended for the general public and that were in charge of disseminating important health information about the COVID-19 pandemic in 2020. The resulting extent of text difficulty was greater than the degree of text difficulty that is normally based on levels for senior primary schools for health promotion materials. The target audience (public or professional) had no bearing on this. Effective engagement with the text posted on both sites required reading at the senior secondary level.

RESEARCH DESIGN

This study targeted datasets of "Open Data Pakistan". Datasets readability was evaluated by a measure (Flesch-Kincaid Grade Level and Flesch Reading Ease score). 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 simpler it is to read and comprehend the information/data, and the harder to understand the content the lower the score. 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). Further, 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 et al. 2018; Risoldi et al. 2012).

The readability findings of 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 values of 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, were output by an online readability checking tool using the Open Data Pakistan URL (https://opendata.com.pk/). To evaluate the open data portal's standard readability, the average scores of each formula were calculated.

The researchers evaluated how many datasets exist, showcases, and the status of open data to check the availability and accessibility by visiting the website in May 2023. Further, each showcase was explored to respond to the study's research question.

RESULTS AND DISCUSSION

Current Status of the Open Datasets

The study found 837 datasets (818 secondary research and 19 primary research). The datasets belong to 14 categories including Public Safety, Economy & Finance, Health, Education, etc. from different locations of Pakistan such as KPK, Sindh, Baluchistan, Islamabad, Lahore Gilgit Baltistan, and Azam & Jammu Kashmir. Datasets are available in different formats i.e. CSV (439), XLSX (198), URL (19), XLS (14) DOCX (5), sav (2), DO (1) DTA (1), ZIP (50), PDF (173), XLS (14) HTML (1) PNG (1) PPT (1) and RAR (1). Additionally, 13 showcases were found on sports, health, gender, attacks, environment, climate, crime reporting, consumer, residential electricity consumption dataset, brain drain, COVID-19, and suicide bombings. Users can access the full datasets page by clicking on the launch website/preview or directly clicking on

the download button. There is a facility to view, download, share, add to favorites, and even nominate for Viz of the day for a showcase. Any user may view the number of views of a showcase. However, we cannot see the number of downloads and shares. Further, we may view comments if a data set has been commented on by users. These datasets are made available to the public for free, allowing individuals, researchers, and organizations to access and analyze the data. The availability of open data promotes 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.

These findings show that Open Data Pakistan (ODP) has a good number of datasets from different fields of life. It also indicated that all Pakistan units represent datasets of ODP. Datasets are strengthened by the different formats, i.e. CSV, XLSX, URL, XLS, DOCX, sav. Users can easily access the complete datasets, which they can explore, download, share, and even submit as candidates for the Viz of the Day display. Open Data Pakistan offers the "connect" option for comments, complaints, and innovative ideas relating to open datasets. Open data accessibility encourages accountability, openness, and fact-based decision-making in Pakistan.

Readability Findings

A key mathematical formula for predicting the reader's level of comprehension of the written piece is the readability measure. An evaluation of readability describes how simple it is to give the reader content. A text document's readability score affects both the content's accessibility and reading speed. When a document's text is inadequate, it makes the readers understand it. 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 in 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 information is easier to read and understand the higher the score, and the harder to understand the content the lower the score. 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.

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

The readability score of FKRE is 47.4, which indicates that OPD's content is hard to read for the general citizens. Literature also supported this result and indicated that the FKRE score was above the average cut of value (Akgül 2019, 2022; Ismail et al. 2019; Serry et al. 2022).

(b) Flesch–Kincaid Grade Level (FKGL)

The essential metrics used by the Flesch-Kincaid Grade Level (FKG) 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 treats any grade value above 12 as being equal to grade value 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 ODP is 7.8 indicating that the data readability is good. It provides evidence that readability is easily comprehended by the general public. The findings supported by the Ismail et al. (2019) and Risoldi et al. (2012) studies, however, some studies found FKGL score difficult to read for the average grade level public (Akgül 2019, 2022; Serry et al. 2022).

(c) Gunning Fog Index (GFOG)

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). Table 3 shows the GFI score and grade level.

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 in 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 abbreviation. A readability test that pays close attention to word length to determine the amount of knowledge necessary for someone to read and comprehend the written word. 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 OPD 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 et al. (2012) study showed SMOG score was insignificant for the readers.

(e) Coleman Liau Index (ClI)

Instead of syllables per word and sentence length, it is based on letters. To grasp the text, it also employs a US grade-based system. 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 (Coleman and Liau, 1975). Table 3 shows the details of the Coleman Liau Index (adopted https://clickhelp.com/).

The Coleman Liau Index score for the ODP 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)

It is derived from 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 demonstrates that if ARI outputs the number 10, that corresponds to a high school student, who is 15 to 16 years old; a number 3 indicates that third graders, who are 8 to 9 years old, should be able to understand the material. The ARI output score for ODP is 5.3. It indicated that students from grade 5 and above (ages 10-11 years) can comprehend the text. This also suggested that the citizens having 5th-grade education can read the datasets of ODP. The findings are also supported by the literature (Ismail et al. 2019).

Summary of the Readability Findings

The summary of the readability evaluation of ODP is shown in Table 2. The readability analysis of ODP shows that FKRE is 47.4, indicating that the content of ODP is challenging for the general public to read. Further, the FKGL score for Pakistan's open data is 7.8, which indicates

that the readability of the data is good. It demonstrates that readability is easily understood by the general audience. Moreover, according to the results of the average GFI readability evaluation, ODP has an average GFI score of 6.2. Based on word metrics, this GFI score revealed that the readability of ODP is simple to understand even for citizens with a six-grade education level. Open Data Pakistan has an SMOG index score of 6.2. It shows that secondary readers (those between the ages of 13 and 15) can easily understand the written text. Open Data Pakistan has a score of 14.8 on the Coleman Liau Index. This shows that the text is too hard for the majority of the readers. The ARI output score is 5.3. It suggested that students in grades 5 and up (about 10 to 11 years old) can understand the text.

Table 2: Summary of the Readability Findings

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

The readability tool also provided an overall score of the readability of the websites. The overall average readability score of the ODP is 47.4 out of 100. The 100 is the highest readability score. The readability tool output score of the ODP shows a very low score indicating readability of the content is difficult to comprehend for the average citizen. On a scale of 0 to 100, at least a 60 score is an acceptable readability recommendation for web content (Flesch 1948). A text with a reading score of 60 to 70 is similar to a grade level of 8 to 9, thus 13 to 15-year-olds should be able to understand it.

These findings suggested that "Open Data Pakistan" may call for a higher level of reading comprehension, making it appropriate for readers with advanced degrees or specialized subject expertise. For readers who are less knowledgeable about the subject, it could be advantageous to simplify the writing or include more explanations. Government operations and decision-making processes can be better understood as a result of open data, which makes it possible for people, researchers, and businesses to access and analyze government data. Governments can streamline their operations and enhance service delivery by increasing data-driven policy formation and evidence-based decision-making. Open data also enables the development of cutting-edge services and products that benefit society as a whole.

CONCLUSION

This study evaluated the availability, accessibility, and readability of datasets, showcasing, and status in the "*Open Data Pakistan*" (ODP) portal. It evaluated the ODP portal's capability for comments, complaints, feedback, and ideas. The study found 837 datasets in the ODP portal along with 14 categories till May 2023. These datasets are available in different formats. Citizens may access full datasets to view, share, and even download them. Moreover, the "Connect" tab is available on the home page for feedback, complaints, and suggestions regarding the ODP portal. Additionally, the readability was checked by the readability checker tool online. The overall readability statistics showed that ODP (https://opendata.com.pk/) has a 47.4% ease of reading which is low and indicates hard to read.

It ought to be simple enough for 13 to 14-year-olds to understand. The availability and accessibility of open data in Pakistan for the citizens are satisfactory however, the readability of the website is not good (47%). It is suggested that the proportion of open data portals in the difficult-to-read category be reduced, increasing the accessibility of the information resources by including easy-to-read phrases in web material.

This study is limited to assessing the readability analysis of ODP. It did not cover the content analysis or usability analysis of ODP. Further, this study used the Flesch-Kincaid Grade Level and Flesch Reading Ease Score to check the readability of Open Data Pakistan. Although it is a well-known scale used to measure the readability of open government portals. However, this online tool for the readability evaluation of ODP can be a limitation of this study.

The study might offer insight into the success of these activities in reaching and involving the public by assessing the readability of open data in Pakistan. It can provide knowledge about how open data can be made more available to promote active participation and give citizens the capacity to make wise decisions. The study also has ramifications for stakeholders and policymakers on the readability, accessibility, and availability issues with open data, emphasizing the significance of taking readability into account in data publication policies and guidelines. It will open new avenues to discuss readability, accessibility, and usability of open data portals in Pakistan that lead to set standards, guidelines, or best practices.

REFERENCES

Abanikannda, D. T., Oke, O. O., Oladipo, F. O., and Ayoola, T. A. 2017. Accessibility and Utilization Pattern of Open Data among Agricultural Science Researchers in South Western Nigeria. *Futo Journal Series (FUTOJNLS).* Vol. 3, no. 2: 9 – 15

Akgül, Y. 2019. The accessibility, usability, quality, and readability of Turkish state and local government websites is 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,* 1-16.

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. doi:10.1016/j.giq.2008.01.004

Clemmensen, T., and Katre, D. 2012. Adapting e-gov usability evaluation to cultural contexts. *In Usability in government systems,* 331-344. doi: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.

Dawes, S. S. 2010. Stewardship and usefulness: Policy principles for information-based transparency. *Government information quarterly,* Vol. 27, no. 4: 377-383.

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. .

Gill, M., and Corbett, J. 2017. Downscaling: Understanding the influence of open data initiatives in smaller and mid-sized cities in British Columbia, Canada. The Canadian *Geographer/Le Géographe Canadien,* Vol. 61, no. 3: 346-359.

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

Henry SL. 2018. *Referencing and linking to WAI guidelines and technical documents.* Available from: 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 the credibility of e-government: Users' perspective. In Human Interface and the Management of Information. *Information and Interaction: Symposium on Human Interface 2009, Held as part of HCI International 2009*, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part II (pp. 26-35). Springer Berlin Heidelberg.

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., & 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. Naval Technical Training Command Millington TN Research Branch. https://doi.org/10.1007/ s10209-022-00909-4 224.

Kitchin, R. 2014. The data revolution: Big data, open data, data infrastructures, and their consequences. *The Data Revolution,* 1-240.

Kozievitch, N. P., Fonseca, K. V., Rosa, M. D. O., Berardi, R., Gutierrez, M. B., Belizario, M., ... and Liu, E. 2022. September). Assessment of Open Data Portals: A Brazilian case study. *In 2022 IEEE International Smart Cities Conference (ISC2)*, 1-7. IEEE.

Laksono, A. D., Wulandari, R. D., and Soedirham, O. 2019. Urban and rural disparities in hospital utilization among Indonesian adults, Iranian. *Journal of Public Health*, Vol. 48, no. 2: 247–255. https://doi.org/10.18502/ijph.v48i2.819

Lazar, J., Beavan, P., Brown, J., Coffey, D., Nolf, B., Poole, R., ... and Wenger, B. 2010. *Investigating the accessibility of state government websites in Maryland. In Designing Inclusive Interactions: Inclusive Interactions Between People and Products in Their Contexts of Use* (pp. 69-78). London: Springer.

Gurstein, M. 2013. *Should "open government data" be a product or a service (and why does it matter?)* Gurstein's Community Informatics. Available at: https://gurstein.wordpress.com/2013/02/03/is-open-government-data-a-product-or-a-service-and-why-does-it-matter/

Máchová, R., Hub, M., and Lnenicka, M. 2018. Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. *Aslib Journal of Information Management,* Vol. 70, no. 3: 252-268.

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

Nikiforova, A., and McBride, K. 2021. Open government data portal usability: A user-centered usability analysis of 41 open government data portals. *Telematics and Informatics,* Vol. 58, 101539.

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.

Olalere, A., and Lazar, J. 2011. Accessibility of US federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly,* Vol. 28, no. 3: 303-309.

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.

Reis, A., Barroso, J., and Gonçalves, R. 2013. Supporting accessibility in higher education information systems. In Universal Access in Human-Computer Interaction. Applications and Services for Quality of Life: *7th International Conference*, UAHCI

2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part III 7 (pp. 250-255). Springer Berlin Heidelberg.

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.

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.

Sheoran, S., Mohanasundaram, S., Kasilingam, R., and Vij, S. 2023. Usability and Accessibility of Open Government Data Portals of Countries Worldwide: An Application of TOPSIS and Entropy Weight Method. *International Journal of Electronic Government Research (IJEGR),* Vol. 19, no. 1: 1-25.

Verkijika, S. F., and De Wet, L. 2017. Determining the accessibility of e-government websites in Sub-Saharan Africa against WCAG 2.0 standard. *International Journal of Electronic Government Research (IJEGR),* Vol. 13, no. 1: 52-68.

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

Wibowo, B. S., Aditya, R. B., and Harianto, T. F. 2021. Harnessing open data and technology for the study of accessibility: The case of Indonesia's capital site candidate. *Spatium*, 46-53.

Yeung, A. W. K., Wochele-Thoma, T., Eibensteiner, F., Klager, E., Hribersek, M., Parvanov, E. D., ... & 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.

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., and Dwivedi, Y. K. 2015. Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government information quarterly,* Vol. 32, no. 4: 429-440.

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.