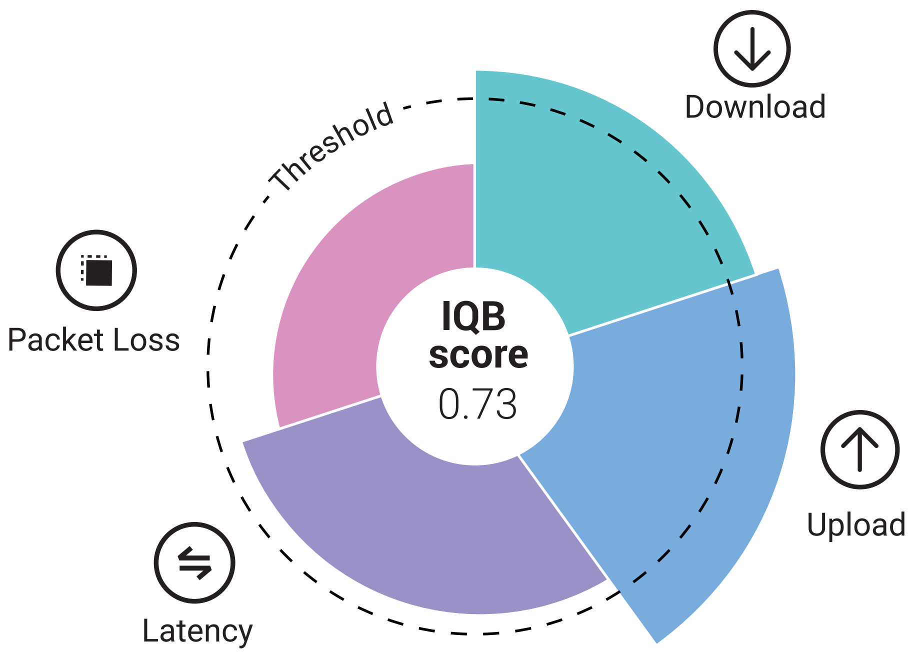
Internet Quality Barometer Framework

Redefining Internet quality beyond speed



Measurement Lab

CS&S Code for Science & Society

The IQB project is funded by

Internet Society Foundation

Index

Executive Summary

Introduction and Goals

The Internet Quality Barometer (IQB), an initiative led by Measurement Lab and funded by the Internet Society Foundation's Research Grant program, seeks to redefine Internet quality beyond "speed." This public report introduces a holistic framework that will be the basis for an IQB tool that will provide stakeholders with actionable insights that support smarter policies and a more equitable Internet. The goals of IQB are:

Shift the conversation around Internet Quality beyond speed

We want to shift the focus of policymakers and advocates beyond speed as the only measure of Internet quality and spark conversations about a broader set of meaningful metrics.

Help decision-makers make sense of the data

With so much data out there, it's easy to feel overwhelmed. IQB aims to simplify the conversation by giving decision-makers the information framework they need to turn information into smart, impactful decision-making strategies.

Empower users to make more informed decisions about their Internet

By providing clear, accessible insights, IQB helps users understand their Internet experience and advocate for better service and policies.

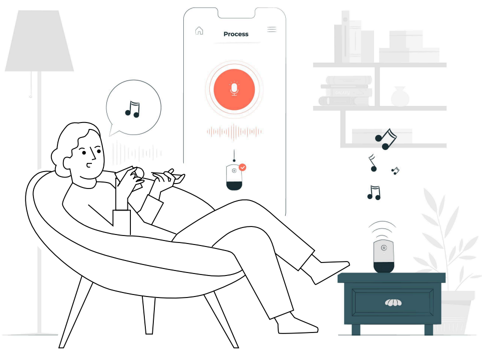
Use existing, openly available datasets as complementary sources

Rather than try and standardize upon one measurement methodology, IQB aims to make use of the multiple datasets that exist by treating them as complementary pieces of a larger puzzle.

Advocate for the collection of more nuanced metrics

That said, while we're helping policymakers understand the data we already have, we're also pushing for new metrics that dig deeper and capture more nuanced complexities of Internet quality.

At Measurement Lab (M-Lab), **we believe that better data will lead to a better Internet.** By accomplishing the goals above, IQB will help foster a more transparent and accessible understanding of Internet quality, and provide the framework through which an improved Internet can be developed.



Gaming and audio streaming are two common use cases for modern Internet users.

The IQB Framework

To create the IQB framework, M-Lab engaged with more than 60 experts across various fields, including academic network research, public policy, digital inclusion advocacy, Internet service provision, speed test data analysis, content provision, and other related domains from November 2023 to March 2025.

From the outset, we recognized the importance of creating a framework that is **accessible to high-level decision-makers** while also garnering **buy-in from experts**. The IQB framework takes inspiration from the credit score—a measure of an individual's "creditworthiness" used by financial institutions—and the Nutri-Score, a nutritional rating system that evaluates the nutritional value of food products. These examples illustrate how a single score can provide a generalized or approximate assessment while being grounded in expert consensus about the key factors and formula used to calculate it. Such frameworks also allow users to explore the underlying formula for greater transparency and provide experts with a platform to contribute feedback, suggestions, and critiques. This iterative process is akin to how standards evolve within organizations like the IETF or other Internet governance bodies.

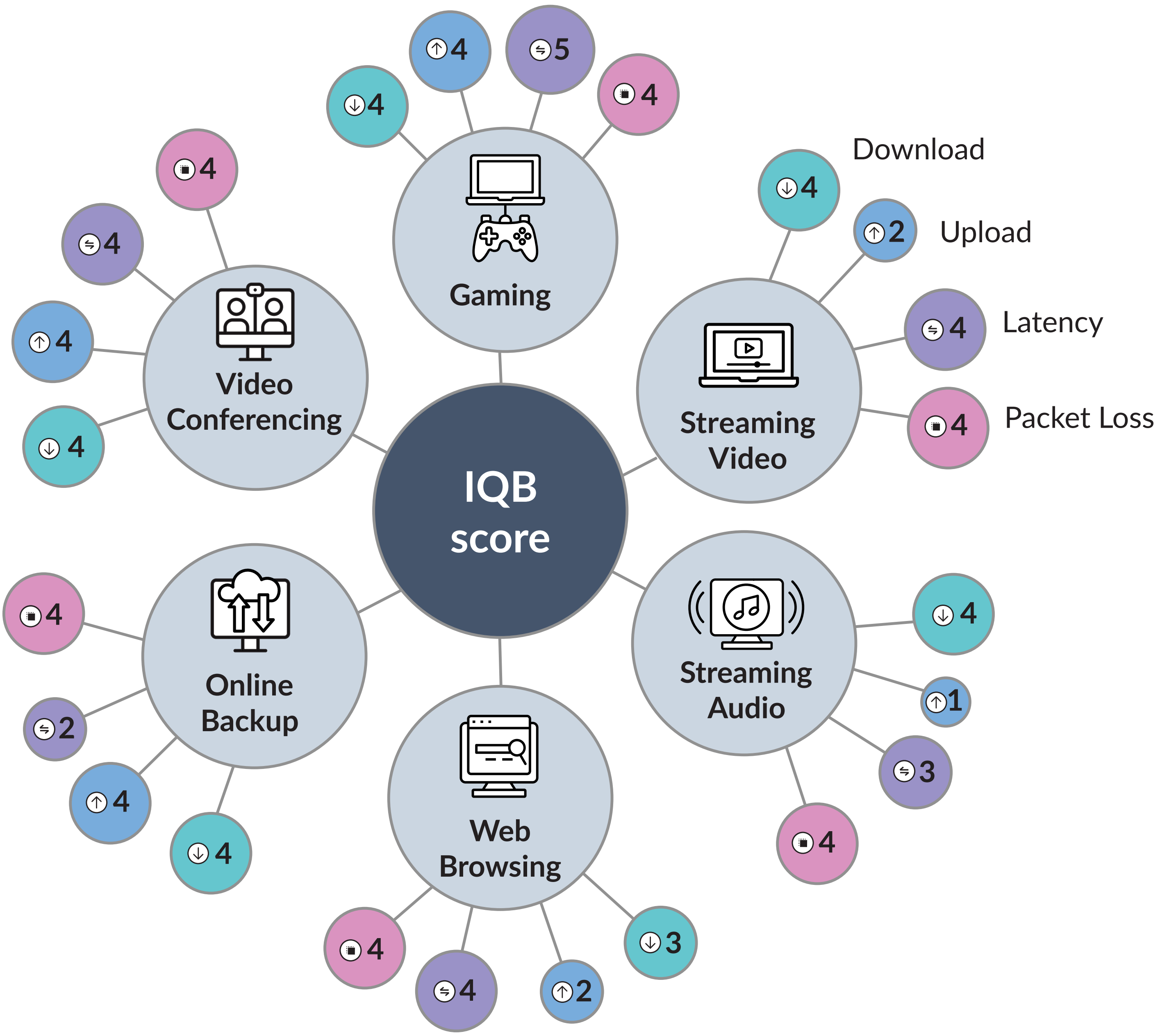
The IQB framework is organized into three tiers: **use cases, network requirements,** and **datasets**. The tiered structure bridges different levels of abstraction. While users and decision-makers rarely think of Internet quality in terms of metrics like throughput, latency, or packet loss, they understand it through what the Internet enables them to do.

> **The use cases tier** reflects the macro-level by identifying activities users should be able to perform online to have a high or minimum-quality experience.

> **The network requirements tier** translates each use case into specific technical needs. For example, what network conditions are necessary for a high-quality video streaming experience? This layer highlights nuances often overlooked in speed tests, such as the differing importance of throughput and latency depending on the use case. Throughput may be critical for downloading large files, while latency is essential for video conferencing.

The datasets tier maps these network requirements to openly available datasets. For instance, if video streaming requires at least 100 Mb/s download speed, the datasets layer identifies open Internet measurement datasets that measure this and determines whether or not the results show evidence of meeting those requirements. Metrics may vary in relevance depending on the use case—e.g., Dataset A may better capture latency for video conferencing than Dataset B. To account for such variations, datasets are weighted based on their applicability to specific use cases.

Each tier produces a value which is aggregated at the top tier to produce an Internet Quality Score. The full report has details regarding the aggregate formula.



This figure illustrates how network requirements contribute to use cases, and how the use cases contribute to the IQB score.

The Future of IQB

This report reflects the first stage of the IQB initiative to redefine Internet quality beyond speed. In the first stage, the primary goal was to lay a strong foundation for future iterations, tools and applications of the IQB framework that will help ensure it can last the test of time. **Prospective next stages of IQB would involve creating a data collection tool, a global IQB pipeline and dashboard, establishing a governance structure, and potentially creating new measurement methodologies.**

As Internet use continues to evolve, so too must the ways we measure and advocate for high-quality connectivity. The IQB framework is designed to be iterative, continuously refined based on new insights and advancements in measurement methodologies. By fostering transparency, inclusivity, and expert-driven dialogue, IQB not only enhances our understanding of Internet quality but also equips decision-makers with actionable intelligence to drive positive change. Ultimately, by shifting the conversation away from a single dominant metric and toward a broader, more nuanced understanding of performance, IQB helps ensure that Internet quality keeps pace with the growing demands of users worldwide.

Introduction

As the Internet has grown, so too have our expectations of its performance. With the rise of IoT ecosystems, real-time interactive applications like video conferencing and gaming, and the increasing size of web pages, the demands placed on Internet connections are more complex than ever (*Fig 1*).

A connection considered high-quality in 2000—or even 2010—would struggle to support the most popular use cases of 2024. Yet, while the Internet itself has evolved, our mainstream understanding of how to measure its quality has failed to keep pace. The Internet Quality Barometer (IQB), an initiative led by M-Lab and funded by the Internet Society Foundation's Research Grant program, seeks to redefine Internet quality beyond "speed."

The Supremacy of Speed

For decades, "speed"—typically defined as throughput or bandwidth—has been the dominant metric for assessing Internet quality. Its appeal lies in its simplistic concept; the faster data can move from point A to point B, the better we expect the performance to be. However, this narrow focus on speed and the capacity of a connection overlooks the growing complexity of modern Internet use.

Internet Service Providers (ISPs) have reinforced this oversimplified association between performance and speed by marketing subscription plans based on advertised speeds. Speed test providers have also promoted the metric by promoting easy-to-use tools that present as the most intuitive way to measure

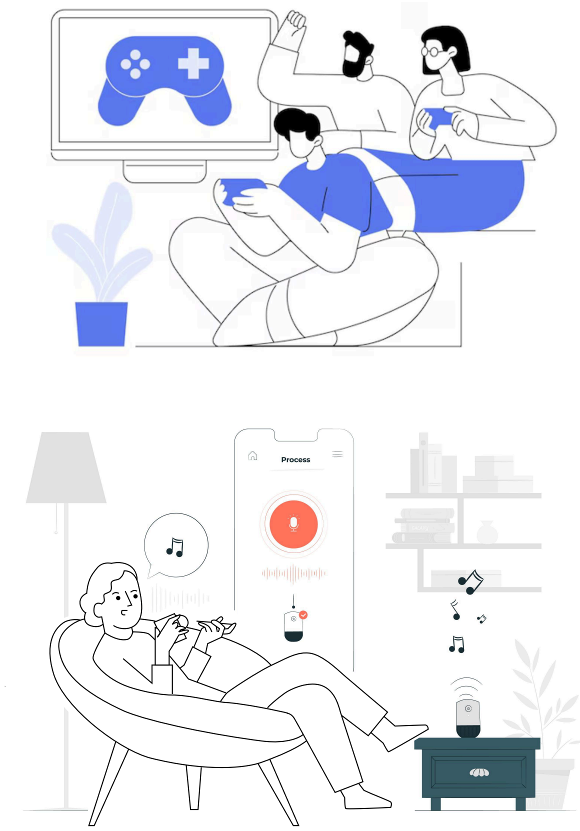


Fig 1. Gaming and audio streaming are two common use cases for modern Internet users.

Internet quality and regulators have further cemented the use of the term by putting terms such as broadband availability in terms of throughput. These strategies have ingrained the notion that faster is the only characteristic that equals better in the minds of consumers, policymakers, and advocates alike. As a result, when decisions are made about improving Internet access, speed is often the primary—and sometimes the only—metric used to measure quality.

For example, the U.S. Federal Communications Commission (FCC) has long relied on ISP-reported data about "available" or "advertised" speeds to track broadband availability. While updates to the National Broadband Map have addressed some shortcomings, the map is still used cautiously due to the inherent limitations of self-reported data. In response, advocates, policymakers, and citizen scientists frequently turn to speed test data, which attempt to provide a closer reflection of users' real-world experiences.

The Challenges of Speed Tests

Speed tests originated during the early days of the commercial Internet and were designed to measure throughput—the bandwidth capacity of a specific connection. At the time, this made sense: the Internet was primarily used for uploading and downloading large files or web pages. However, as the Internet's use cases expanded, so too did the need for more sophisticated metrics, yet the fundamental design of speed tests has remained largely unchanged.

Several factors influence the results of a speed test, including the protocol used (e.g., TCP or UDP), the testing device (e.g., a smartphone or router), the location and selection method of the server, the duration or volume of data transmitted, and the algorithms employed. Each of these elements affects how speed is measured and interpreted, creating inconsistencies between tests.

Popular speed test platforms, such as Ookla, Cloudflare, Fast.com, and Measurement Lab's Network Diagnostic Tool (NDT), differ in their methodologies and goals. For instance:

- Ookla measures speed by opening multiple TCP connections.
- NDT focuses on a single TCP stream for its tests.
- Cloudflare evaluates speed within its own network, emphasizing its application performance.
- Fast.com targets Netflix's servers to gauge streaming performance.

These differences highlight the varying priorities of each provider but also lead to

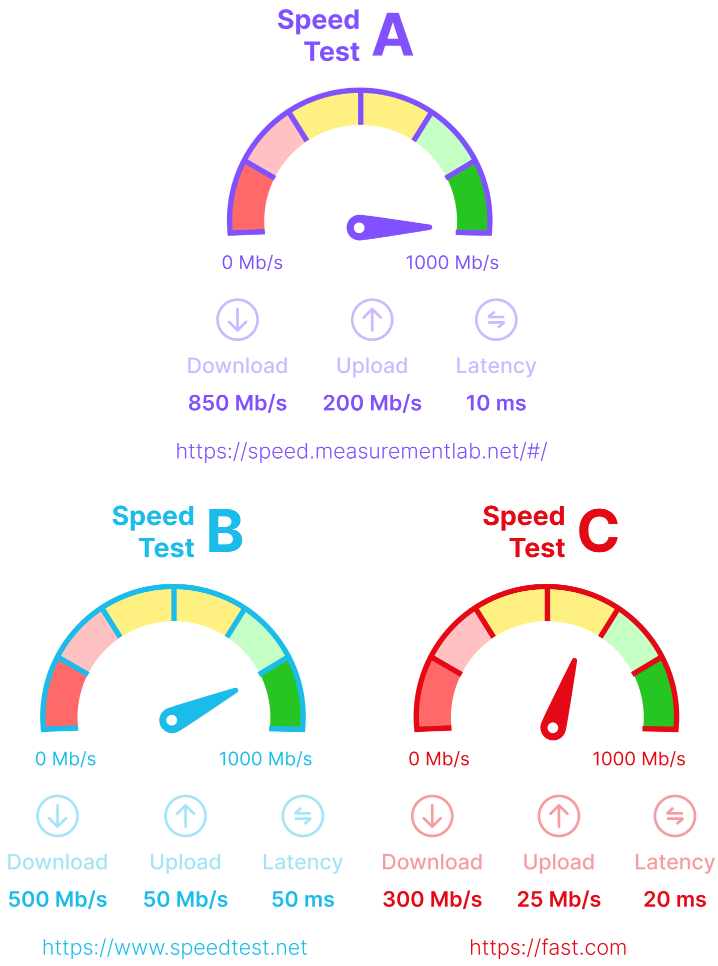


Fig 2. Speed tests often return different results for the same user.

fragmented definitions of "speed." This fragmentation complicates our collective understanding of what constitutes a quality Internet connection and raises questions about how to standardize such measurements.

Multiple Maps, Conflicting Results

For decision-makers, this lack of standardization has led to conflicting datasets and competing broadband maps, making it difficult to form clear strategies. Publicly available data is meant to streamline decision-making, not complicate it, yet policymakers are often left to interpret complex, inconsistent datasets without a unified framework to guide them. While the temptation might be to standardize a single version of "speed," such an approach would limit our ability to

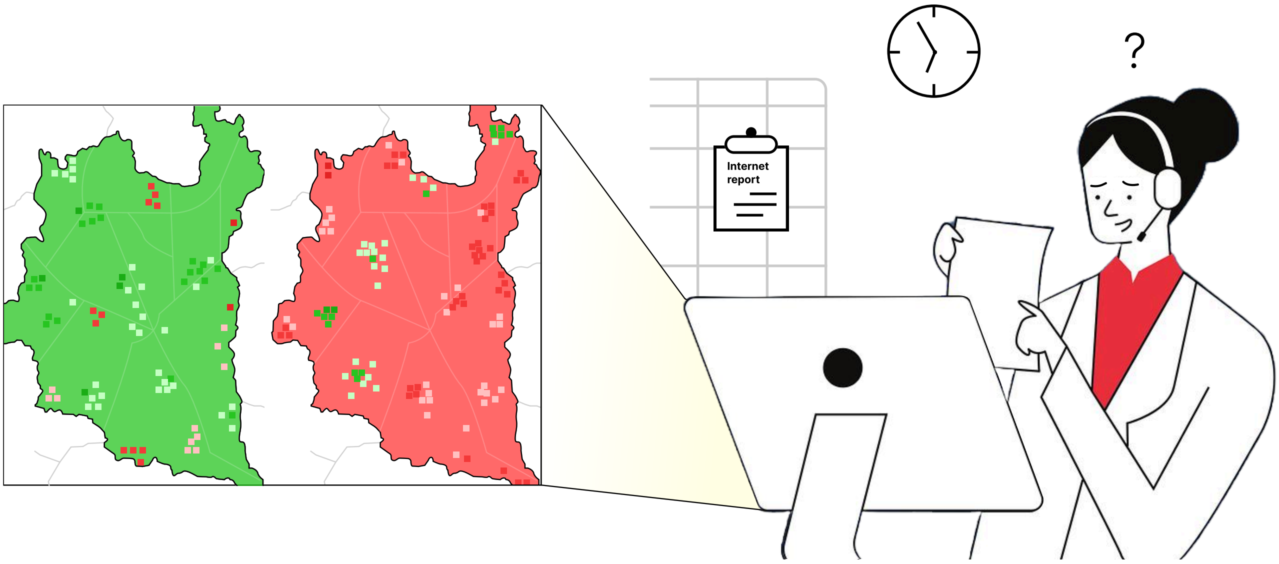


Fig 3. Speed test maps often present varying results, making interpretation a challenge for policymakers.

measure the Internet's multifaceted performance. For example, choosing between Ookla's and NDT's methodologies—or even creating a hybrid, singular approach—would still capture only part of the story. Instead, embracing the discrepancies between these tests, and framing them as complementary perspectives, could provide deeper insights into the strengths and weaknesses of a connection.

Beyond their methodological differences, speed test datasets also suffer from an overemphasis on throughput as the primary indicator of quality. Metrics like latency and packet loss, which are critical to modern applications such as gaming and video streaming, are often underutilized in high-level decision-making tools like broadband maps. This gap stems from a persistent cultural bias equating speed with quality and a lack of frameworks to help decision-makers understand why these additional metrics matter.

Motivation and Goals

The Internet Quality Barometer (IQB) project is driven by the need to redefine

how we measure and understand Internet performance to keep pace with evolving technological demands and user expectations. By moving beyond a narrow focus on speed, the project seeks to provide stakeholders with actionable insights that support smarter policies and a more equitable Internet. Valuable but requires a framework to make that data both accessible and actionable; IQB aim's primary goal is to fill that gap.

Our goals are to:

Shift the conversation around Internet Quality beyond speed

We want to shift the focus of policymakers and advocates beyond speed as the only measure of Internet quality and spark conversations about a broader set of meaningful metrics.

Help decision-makers make sense of the data

With so much data out there, it's easy to feel overwhelmed. IQB aims to simplify the conversation by giving decision-makers the information framework they need to turn information into smart, impactful decision-making strategies.

Empower users to make more informed decisions about their Internet

By providing clear, accessible insights, IQB helps users understand their Internet experience and advocate for better service and policies.

Use existing, openly available datasets as complementary sources

Rather than try and standardize upon one measurement methodology, IQB aims to make use of the multiple datasets that exist by treating them as complementary pieces of a larger puzzle.

Advocate for the collection of more nuanced metrics

That said, while we're helping policymakers understand the data we already have, we're also pushing for new metrics that dig deeper and capture more nuanced complexities of Internet quality.

Ultimately we believe that better data will lead to a better Internet. By accomplishing the goals above, IQB will help foster a more transparent and accessible understanding of Internet quality, and provide the framework through which an improved Internet can be developed.

Background

Starting in November 2023, M-Lab engaged with more than 60 experts across various fields, including academic network research, public policy, digital inclusion advocacy, Internet service provision, speed test data analysis, content provision, and other related domains. These discussions aimed to identify key elements for a framework to measure Internet quality, particularly to inform decisions about investments to improve Internet infrastructure in specific regions.

In the following section, we outline the inspiration behind the framework, our process for gathering feedback, and the known challenges and limitations.

Audience and Stakeholders

IQB is designed for high-level decision-makers like policymakers, advocates and other individuals or organizations with the influence and interest to improve the Internet using data. These are the people who can take insights and turn them into impactful actions (*Fig 3*).

Our initiative also seeks guidance and approval from stakeholders, which include our audience—policymakers and decision-makers—as well as network measurement experts from policy, advocacy, industry, and academic research, with an emphasis on those who provide data that could be included in the framework.

In addition to being a framework, IQB is also a conversational mechanism through which different areas and levels of expertise can mediate their discussions about Internet quality. The language and references that a policy maker uses to describe the characteristics of a region's Internet quality are likely different than that of a network operator - IQB provides a translation tool to traverse these levels of abstraction and identify common points of focus and concern.

Inspiration

As outlined in the Background section, while it might seem appealing to resolve the complexity of multiple speed tests by standardizing around a single methodology, we chose to advocate for an approach that embraces the diversity of insights available from using multiple methodologies. We also sought to create a framework that could incorporate a variety of metrics, such as loaded and unloaded latency, or the more advanced aggregate metrics with a more targeted focus such as Apple's Network Responsiveness Measurement or Quality of Attenuation (Broadband Forum TR-452.1). To explore how these complex, technical concepts could work together, we looked to examples like the credit score—a measure of an individual's "creditworthiness" used by financial institutions—and the Nutri-Score, a nutritional rating system that evaluates the nutritional value of food products. These examples illustrate how a single score can provide a generalized or approximate assessment while being grounded in expert consensus about the key factors and formula used to calculate it.

Such frameworks strike a balance between simplicity and complexity by offering an easy-to-understand number or rating to optimize for, while also, as with the Nutri-Score, allowing users to explore the underlying formula for greater transparency. This transparency not only helps users understand how values are calculated but also provides experts with a platform to contribute feedback, suggestions, and critiques. This iterative process is akin to how standards evolve within organizations like the IETF or other Internet governance bodies.

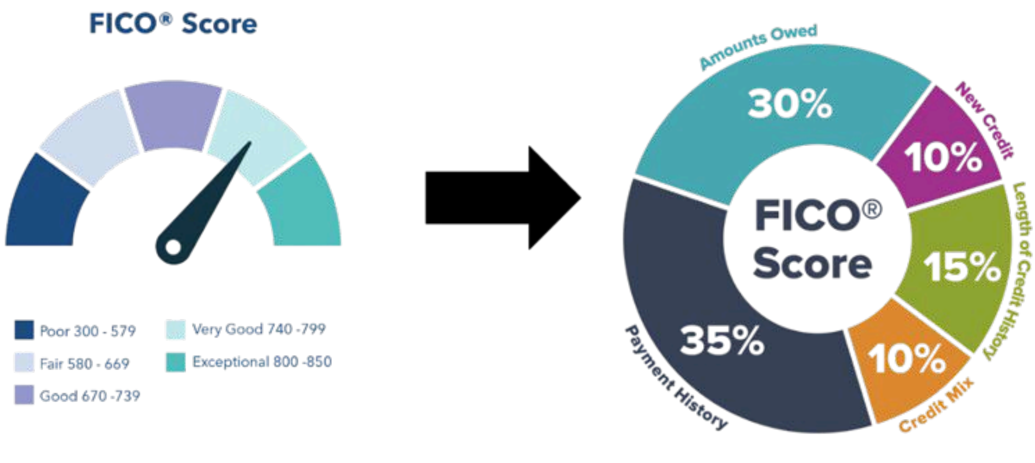


Fig 4. We drew inspiration from the concept of a "credit score," which assesses an individual's creditworthiness based on multiple factors.

Concept

Separating these levels of understanding and accessibility was central to our goal. From the outset, we recognized the importance of creating a framework that is **accessible to high-level decision-makers** while also garnering **buy-in from experts**. The structure of these frameworks as a conceptual inspiration was particularly compelling because they maintain the intuitive simplicity of a single, actionable number.

> *What if we could replace that singular focus on speed? And what should the new metric include?*

One of the current challenges with the dominance of “speed” as a measure is the disproportionate emphasis it places on optimizing networks for bandwidth. What if we could replace that singular focus on speed? And what should the new metric include? Through our research, we posed this question to experts in academia, industry, and policy. Their responses—collected through workshops, interviews, and a comprehensive survey—have informed the first iteration of the formula. As more stakeholders contribute their perspectives, the framework will continue to evolve.

Process

We conducted interviews with stakeholders from institutions including: Meta, University of California Berkeley, University of California Santa Barbara, MIT, Réseaux IP Européens Network Coordination Centre, Rede Nacional de Ensino e Pesquisa, Marconi Society’s National Broadband Mapping Coalition, and Global Digital Inclusion Alliance. These interviews helped inform the preliminary ideas of the framework and produced our first iteration.

We also engaged 50 experts in international workshops in Madrid and Barcelona, Spain, and Washington, D.C. These locations were chosen to optimize the travel costs for regional experts and increase the number of local experts that were able to participate. The half-day workshops were designed to share the goals of the IQB framework and solicit their feedback on its formula and structure. Each stakeholder was shown a prototype of the framework and asked questions about its structure, organization and development process and helped develop the framework further into the more detailed structure that is published in this report.

Finally, we sent out a call for responses to a survey to help define the specifics of the framework, in particular the thresholds for each network quality level and the weighting of each metric per use case. For more information on how the survey results were used, please refer to the Framework description.

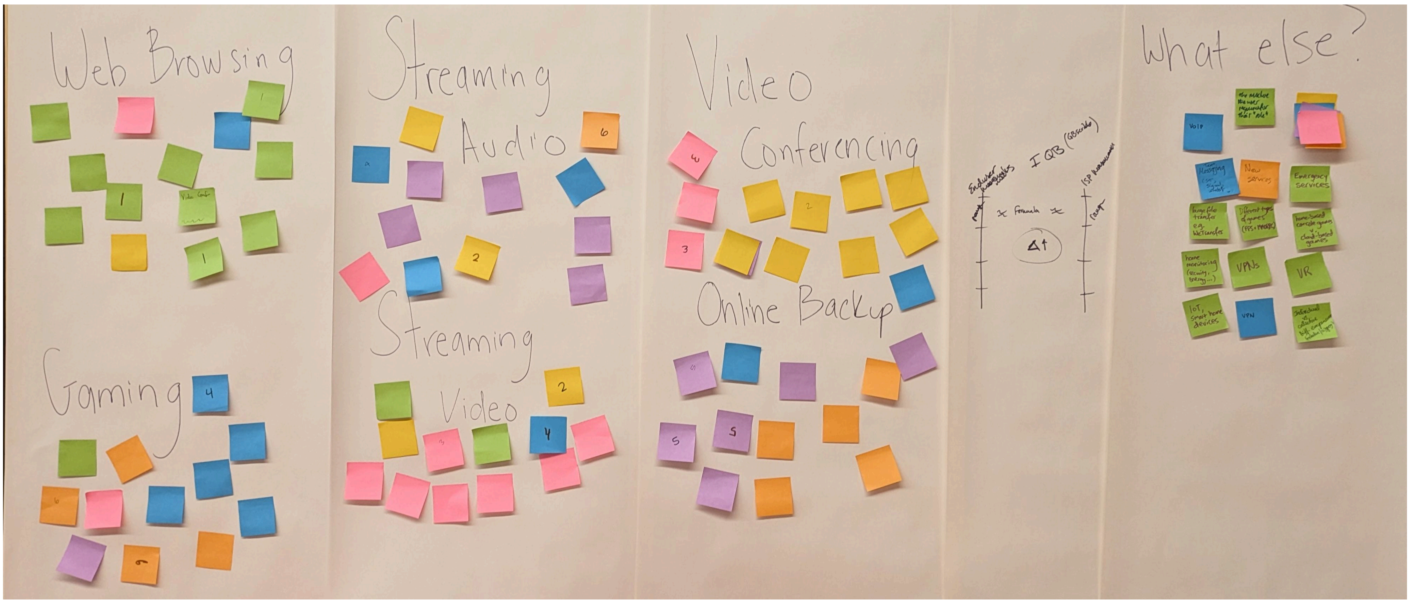


Fig 5. Participants shared input on how to measure each use case.

Fig 6. Participants of the workshop in Barcelona. September 2024.

Fig 7. Participants of the workshop in Barcelona. September 2024.

Framework

Approach

Rather than developing a measurement methodology from scratch, the IQB takes a top-down approach by analyzing and interpreting existing data and methodologies. Our goal was not to create a definitive or perfect representation of Internet quality. Instead, we aimed to develop a high-level aggregation that strikes a balance between being comprehensive and providing actionable insights. To that end, we prioritize generalizations over specificity; IQB is not a finely tuned measurement tool, but instead a sieve to make the coarsely grained insights provided by public data more accessible.

> We prioritize generalizations over specificity.

With this balance in mind, we recognized the need to account to make the variance in quality of measurement data more visible. The IQB framework is built to leverage openly accessible data, which, while valuable, often provides only an approximation of Internet quality due to limitations in methodology, collection techniques, and other factors. Combining multiple datasets and methodologies helps mitigate some of these limitations, but gaps and constraints in the data may still exist. To address this, each dataset incorporated into the framework is given a value (described below as a "weight"), which determines how much it can address the score. With this approach, a dataset can have an impact on an IQB score, while also not solely defining Internet quality, as we often see with speed tests.

Finally, as mentioned earlier, the IQB Framework is an iterative and collaborative project. We acknowledge that perfection won't be achieved immediately (and perhaps not in the first, second, or even third attempt). It will require multiple iterations and input from a diverse range of stakeholders to develop a framework and formula that achieves broad consensus. Even then, it will need to evolve alongside changes in the Internet and our understanding of quality. The proposed process for iteration is outlined in the Iteration section of the report.

Summary

The IQB framework is organized into three tiers: **use cases, network requirements,** and **datasets *(Fig 8)***. The tiered structure bridges different levels of abstraction. While users and decision-makers rarely think of Internet quality in terms of metrics like throughput, latency, or packet loss, they understand it through what the Internet enables them to do.

To reflect this macro-level understanding, the IQB framework begins with a **use cases tier**, identifying activities users should be able to perform online to have a high or minimum-quality experience. In essence, what does Internet quality allow a user to achieve?

Next, the **network requirements tier** translates each use case into specific technical needs. For example, what network conditions are necessary for a high-quality video streaming experience? This layer highlights nuances often

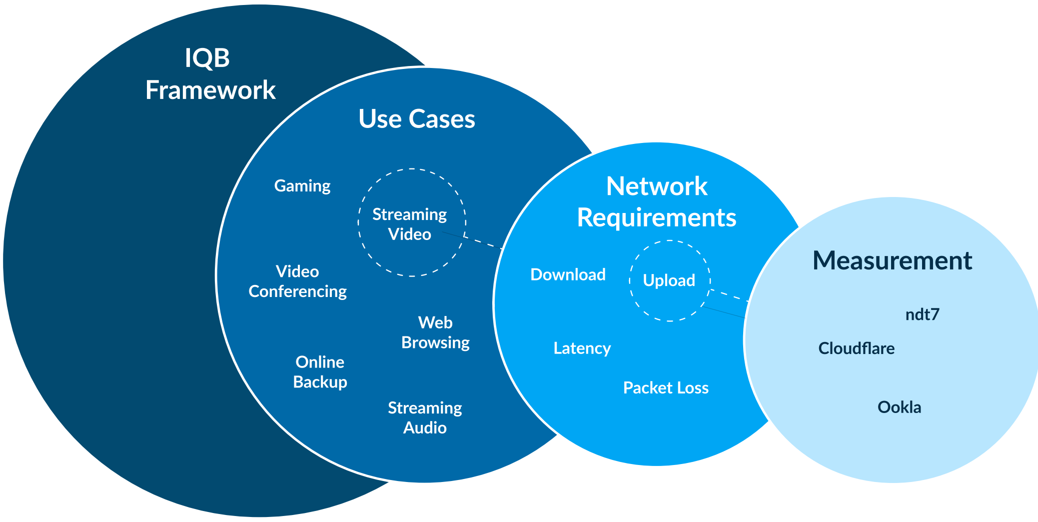


Fig 8. The IQB Framework has 3 tiers: Use Cases, Network Requirements and Measurement.

overlooked in speed tests, such as the differing importance of throughput and latency depending on the use case. High throughput may be critical for downloading large files, while low latency is essential for video conferencing.

Finally, the **datasets tier** maps these network requirements to openly available datasets. For instance, if experts agree that video streaming requires at least 100 Mb/s download throughput, the datasets layer identifies open Internet measurement datasets that measure download throughput and determines whether or not the results show evidence of meeting those requirements. Metrics may vary in relevance depending on the use case—e.g., Dataset A may better capture latency for video conferencing than Dataset B. To account for such variations, datasets are weighted based on their applicability to specific use cases.

The second and third tiers work together to align network requirements with the datasets that assess them. Additionally, some advanced methodologies, such as Apple's Network Responsiveness metric or Domos' Quality of Outcome metric, already integrate network requirements and measurement into composite metrics.

Since they already do the translation of basic metrics into their own composite metrics on their own, they span both the network requirements and datasets layers. When available, they serve as additional datasets that enhance the quality of the score.

The next sections describe each of these layers in more detail and describe our reasoning and methods for how each tier is defined.

> *In simple terms, if you have high-quality Internet, what does that enable you to do online?*

Use Cases

The goal of the use cases tier is to define Internet quality through a user-centric framework. In simple terms, if you have high-quality Internet, what does that enable you to do online?

As a starting point, we referenced Making Broadband Internet Labels Useful and Usable: Preliminary Report on Consumer-Driven Broadband Label Design, which explored ways to enhance the FCC's Broadband Labels by providing consumers with more detailed information about their broadband subscription packages. The paper proposed separate performance and reliability ratings for six use cases: web browsing, streaming audio, streaming video, online backup, video conferencing, and gaming.

Through workshops and interviews, we reached a consensus that these use cases effectively represent what users should be able to do with their Internet connection. Some stakeholders provided feedback that the choices of which specific use cases we chose were somewhat arbitrary and that they were essentially placeholders for more technical concepts such as whether or not an application required interaction or large amounts of data transfer. Others made clear that they felt the categories were too broad and general to provide specific measurement recommendations for.

However, participants ultimately also recognized the importance of the first tier's broader purpose: to articulate the parameters for defining network requirements clearly while also providing relatable, real-world concepts for our primary audience of high-level decision makers. Though technical terms could make the use cases more precise, keeping them abstracted into accessible actions helps ensure their relevance to high-level decision-makers. This generalization allows decision-makers to connect these use cases to their work, while still enabling technical specificity in subsequent tiers of the framework.

Our use case tier is defined in Fig 9 (see next page):

If there are additional use cases that a user of IQB would like to see reflected in the framework or if there are use cases that are less relevant to their region of focus, then they can either create a custom version of the framework for their own purposes or propose a modification to the framework. See the Iteration and Customization section for more information.

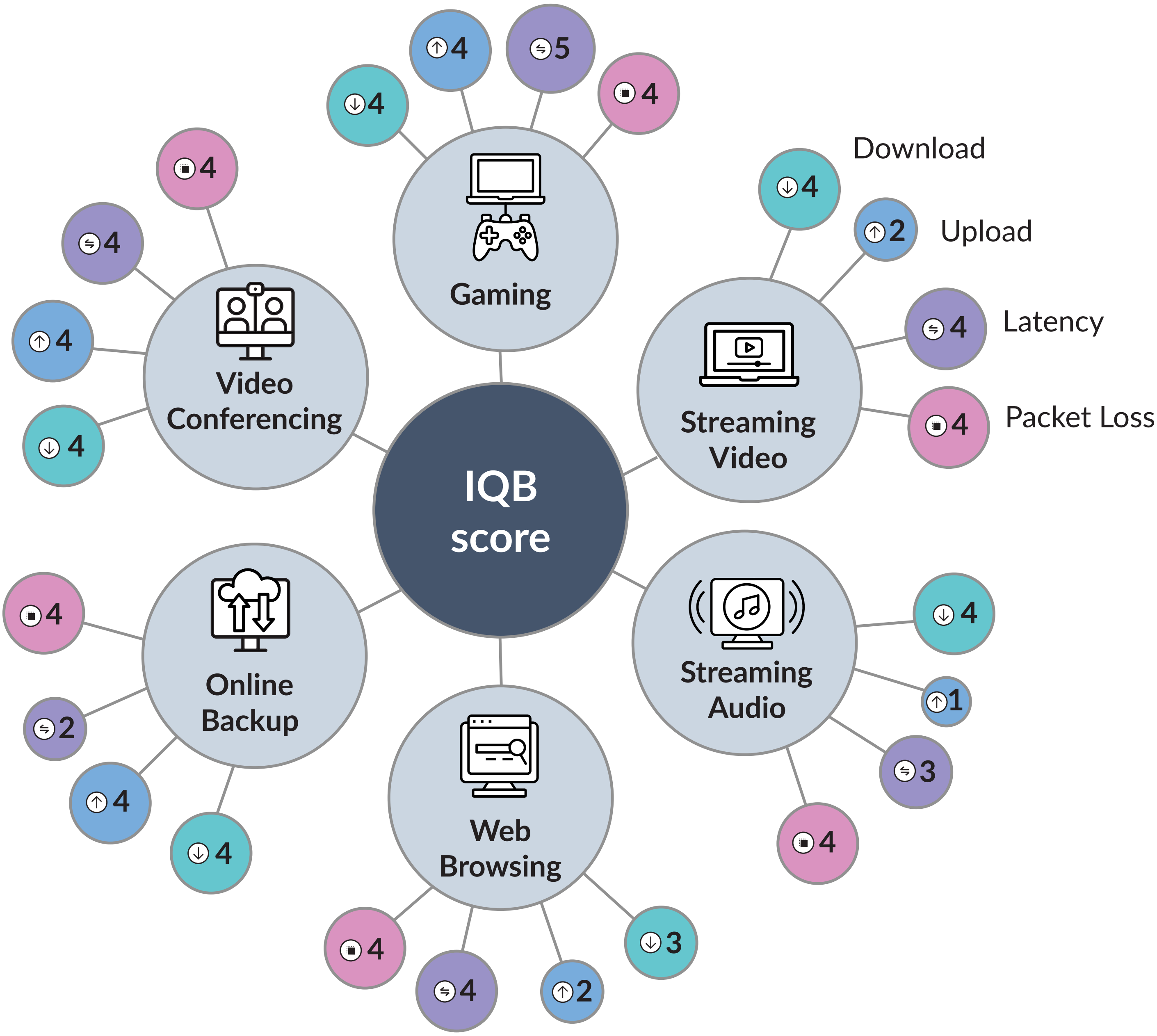


Fig 9. This figure illustrates how network requirements contribute to use cases, and how the use cases contribute to the IQB score.

Network Requirements

The goal of the network requirements layer is to map each use case into more specific technical requirements. In simple terms, “what does a user need from their network to be able to do this use case?”

> *What does a user need from their network to be able to do this use case?*

It does so by aligning each use case with its requirements in terms of basic metrics found in openly available speed test datasets, as well as more advanced, composite metrics that can be used when available. We'll start by reviewing how the thresholds for basic metrics were formulated. For more discussion of the advanced metrics please see below.

Quality thresholds for basic metrics: For each use case, we provide what a user needs in terms of throughput, latency, packet loss and jitter to be considered as having a high quality experience or a minimum quality experience (i.e. can they do it at all?). To define these thresholds, we surveyed a group of experts and through their feedback arrived at the thresholds depicted in *Fig 10* (and presented in detail in the Appendix-I).

For instance, in the Audio Streaming use case, a minimum download and upload throughput of 10 Mb/s is required to achieve acceptable quality. To experience high-quality audio streaming, a throughput of at least 50 Mb/s is recommended. Additionally, to meet the minimum quality standard, latency should not exceed 100 milliseconds, and the packet loss rate should remain below 1%.

The following figure (*Fig 11*) illustrates how quality thresholds for download throughput vary across different use cases, underscoring the motivation behind the IQB framework—namely, the importance of accounting for specific use cases when evaluating overall Internet quality.

Use case	Download Throughput		Upload Throughput		Latency		Packet Loss	
	for min quality	for high quality	for min quality	for high quality	for min quality	for high quality	for min quality	for high quality
Web Browsing	10Mb/s	100Mb/s	10Mb/s	Other	100 ms	50ms	1%	0.5%
Video Streaming	25Mb/s	50-100 Mb/s	10Mb/s	10 Mb/s	100 ms	50ms	1%	0.1%
Video Conferencing	10Mb/s	100Mb/s	25Mb/s	100Mb/s	50 ms	20ms	0.5%	0.1%
Audio Streaming	10Mb/s	50Mb/s	10Mb/s	50Mb/s	100 ms	50ms	1%	0.1%
Online backup	10Mb/s	10Mb/s	25Mb/s	200Mb/s	100 ms	100ms	1%	0.1%
Gaming	10Mb/s	100Mb/s	10Mb/s	Other	100 ms	50ms	1%	0.5%

Fig 10. Network requirements thresholds for minimum and high quality for each use case

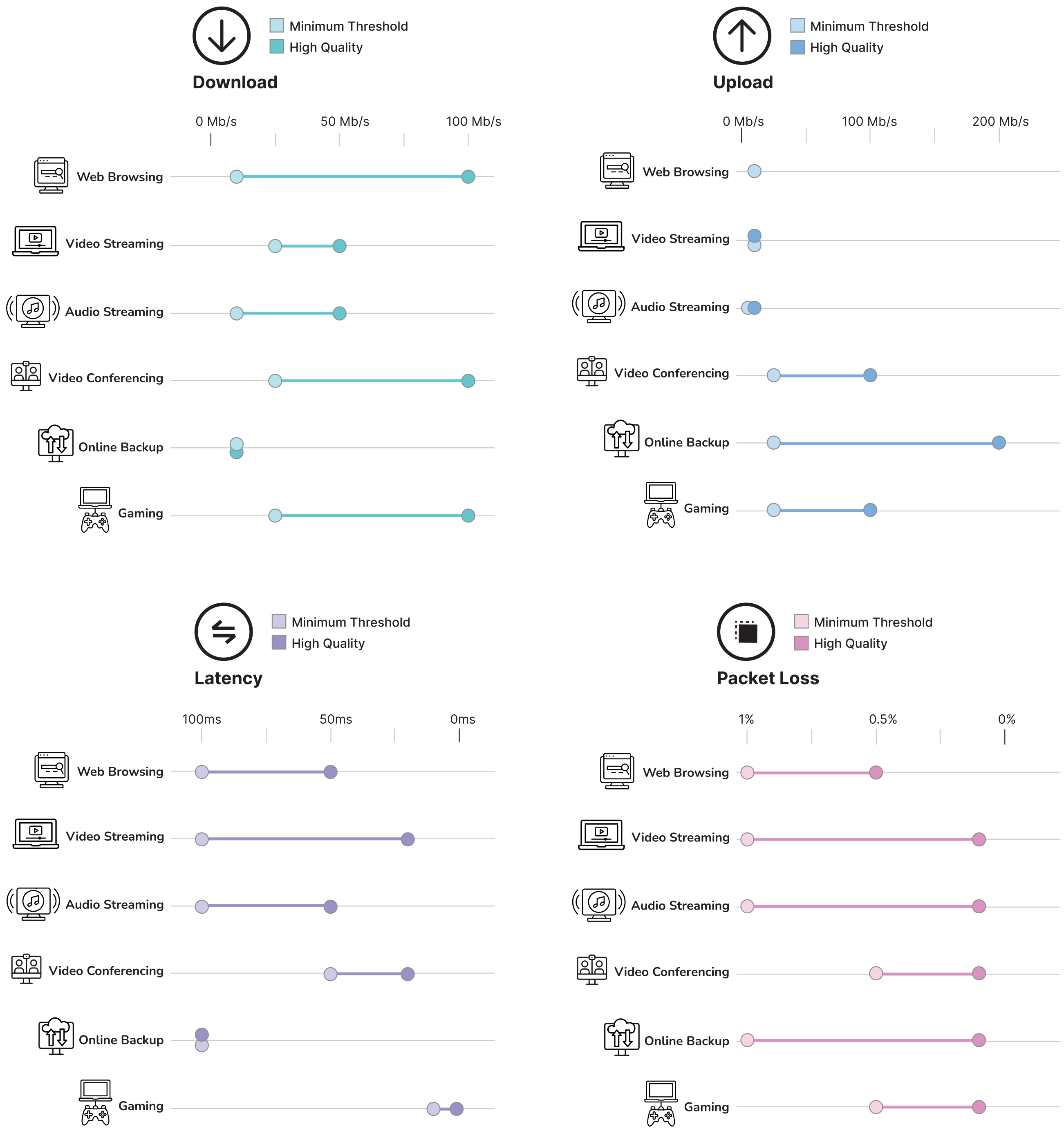


Fig 11. Minimum quality and High quality thresholds for each use case and network requirement. Thresholds can vary significantly across different use cases.

It should be noted that

1. These thresholds only reflect the opinions of 17 experts and
2. None of the answers had full consensus. Therefore, we suggest that these values should be interpreted as a first iteration which can continue to be iterated upon as IQB continues.

Weights per basic metric: One of the current limitations we described about speed test data in our background section is the overemphasis of throughput over other metrics that have more of an impact on the quality of experience. To account for this, we also asked experts to weigh in on how much a given metric matters for a specific use case.

For example, after identifying the minimum latency that was needed for a high quality experience for audio streaming, we asked, “How much does latency matter when performing an online backup? Provide a rating between 1 and 5 with 1 being “ it doesn’t matter at all”, and 5 being “it’s the only metric that matters”. Fig 12 shows the weights assigned to each basic metric for the audio streaming use case. Download throughput and packet loss are identified as the most critical requirements, each with a weight of 4, while upload throughput is considered the least important, with a weight of 1.

The detailed results for the weights per metric per use case are depicted in Fig. 9 (and a detailed table can be found in the Appendix-II).

With these results, we were able to construct a set of network requirements for each use case that can determine whether or not a connection meets the requirements for a high-quality or minimum-quality experience for a given use case. This tier provides more technical specificity to what is required of a connection to be able to support the use cases in the tier above and provides a more user-centric framework through which we can interpret the datasets in the next tier.

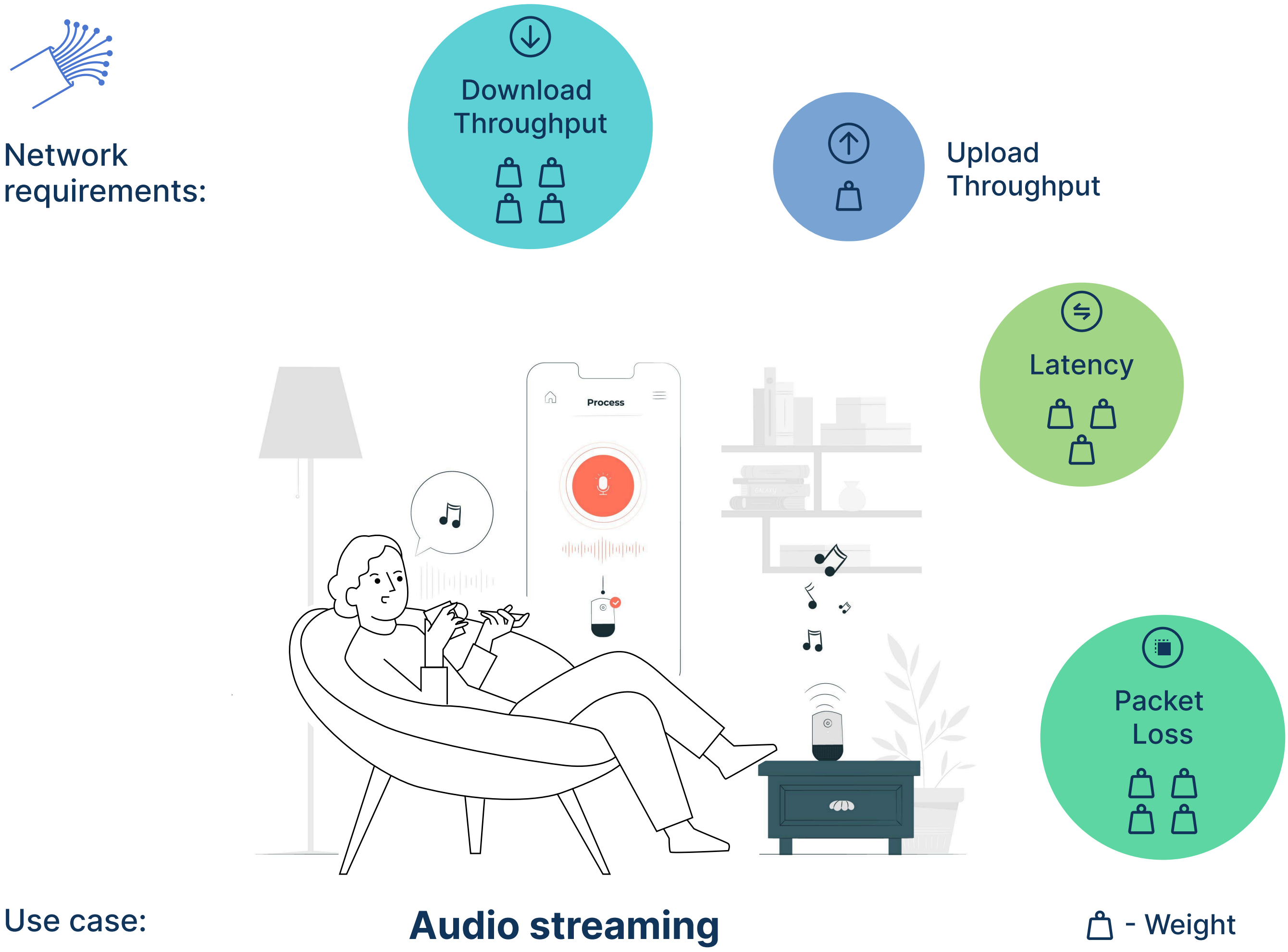


Fig 12. Network conditions necessary for a high-quality **audio streaming** experience.

Datasets

The goal of the third and final tier is to map these network requirements onto openly available datasets. In simple terms, "What does publicly available data suggest about this network requirement?". Three commonly used datasets are M-Lab's NDT and Cloudflare, due to their public availability at the individual test level and Ookla's publication in the aggregate. There are other popular speed test providers such as Fast.com and LibreSpeed, but their results are not made available for open access. Within the following section we'll reference the previously mentioned open datasets as examples of "core datasets" due to their widespread availability in multiple regions, it should be noted that the framework can be adjusted to include any available dataset. It should also be noted that these datasets do not conclusively cover every region and there are known limitations and biases in their data collection methods. We address these limitations and how to communicate them to users in the following section.

The primary benefit of using multiple datasets is using their differences to corroborate each other's insights. For example, because NDT, Ookla and Cloudflare each measure throughput in a fundamentally different way, if they all signal that a connection meets the throughput requirements for gaming, then it is more likely that that connection does meet the requirements. However, if only one of the three datasets signal that the requirements are being met, then it is less likely. The datasets tier provides a framework through which we can interpret the various results these datasets give us in relation to one another by using a weighted average which we call the "per-requirement agreement score". This score reflects the degree to which the available datasets show evidence that a specific network requirement is being met. For more detail, please see the Detailed Description of Formula section.

Weight Values

There are instances where a dataset's relevance to a use case varies. For example, if a dataset is specifically designed to measure interactive applications, its results for interactive use cases such as video conferencing and gaming should be given greater weight. On the other hand, some methodologies may have characteristics that suggest their results should be viewed with caution across all use cases, not just one. For instance, ndt7's measure of packet loss reflects how BBR handles loss rather than the network itself, so its packet loss results should be given less emphasis throughout all use cases. IQB codifies these levels of emphasis through "weights" of each dataset, similar to how the network requirements give distinctive weights to individual metrics.

When new datasets are added, they should be given their own weights to be used in the framework.

Aggregation

IQB uses the 95th percentile of a given dataset to evaluate a given metric. In this context, the 95th percentile is the value below which 95% of the observed measurements fall, which effectively captures the upper bound of a typical user experience while excluding extreme outliers. For example, to assess whether a region meets the network tier's packet loss region meets the network tier's packet loss criteria for high-quality gaming, IQB calculates the 95th percentile of packet loss measurements collected from users in that region. The value is then compared to the predefined threshold.

If an IQB user would prefer to use a different aggregation of a dataset, then they can either create a custom version of the framework for their own purposes or propose a modification to the IQB Global framework. See the Iteration and and Customizing IQB section for more information.

Advanced Datasets

Up until this point we've referenced basic metrics (e.g. throughput, latency and packet loss), but much of the same approach applies to the advanced datasets. The primary difference is that instead of requiring translation between the network requirements tier and the datasets tier, the advanced metrics are both a network requirement and a dataset, and therefore span both tiers. For example, if there has been a statistically relevant amount of Apple's Network Responsiveness data collected in a given region, then we can use the results that the test returns ("Low", "Medium" and "High) to determine if the region has what a use case requires for a given use case e.g. audio streaming requires "medium" network responsiveness.

Advanced datasets are then added to the aggregate formula described above and provide another dataset to compare against the others.

Detailed Description of Formula

To build an IQB score, we start at the "bottom" of the layers with the Datasets tier, and work our way upwards through successive tiers. Below, we outline this process in abstract terms; detailed explanations and examples are provided in the Appendix-III.

Assume that we have four datasets: Datasets A, B, and C return "classic" metrics e.g. download, upload, latency and packet loss. Dataset D returns an "advanced summary" metric e.g. "Low, Medium or High".

- For each dataset, we aggregate its measurements (taking the 95th percentile) and calculate the **binary requirement score** per dataset (**S_A, S_B, S_C, S_D**) that indicate whether or not the threshold for that individual network requirement for a high-quality experience of a particular use case is met.

- We calculate the **requirement agreement score (S_{req})** per network requirement, which is a weighted average (assuming dataset weights **w_A, w_B, w_C, w_D**) of the binary requirement scores. The requirement agreement score indicates the degree to which the available datasets "agree" that a network requirement for a high-quality user experience for a given use case is being met.

- Then we calculate the **use-case score ($S_{use\text{-}case}$)** per use case, which indicates the degree to which the available datasets indicate that users in the region or network will have a high-quality experience of a use case. The use case score is similarly calculated as a weighted average of the requirement agreement scores; where each requirement has a different weight for each use case as discussed earlier.

Finally, we calculate the **Internet quality score, or IQB score, (S_{IQ})** as a weighted average of the use case scores.

Formula cheatsheet

$$S_{req} = \frac{(w_A \times S_A) + (w_B \times S_B) + \dots}{w_A + w_B + \dots}$$

S_{req} - is the overall agreement score for the network requirement e.g. latency or packet loss

$w_{A,}\ w_B$ - are the weights of the individual datasets for the network requirement

$S_{A\,,}\ S_B$ - are the binary requirement values per dataset for the network requirement

$$S_{use-case} = \frac{(w_{DL} \times S_{DL}) + (w_{UL} \times S_{UL}) + (w_L \times S_L) + (w_{PL} \times S_{PL}) + (w_D \times S_D)}{w_{DL} + w_{UL} + w_L + w_{PL} + w_D}$$

$S_{use\text{-}case}$ - Overall score for use case

S_{DL} - *download*

S_{UL} - *upload*

S_L - *latency*

S_{PL} - *packet loss*

S_D - *dataset*

S_{req} - Overall agreement score for network requirement

$$S_{IQ} = \frac{(w_{VS} \times S_{VS}) + (w_{AS} \times S_{AS}) + (w_{VC} \times S_{VC}) + (w_{GM} \times S_{GM}) + (w_{OB} \times S_{OB}) + (w_{WB} \times S_{WB})}{w_{VS} + w_{AS} + w_{VC} + w_{GM} + w_{OB} + w_{WB}}$$

S_{IQ} - Internet quality score

S_{VS} - *video streaming*

S_{AS} - *audio streaming*

S_{VC} - *video conferencing*

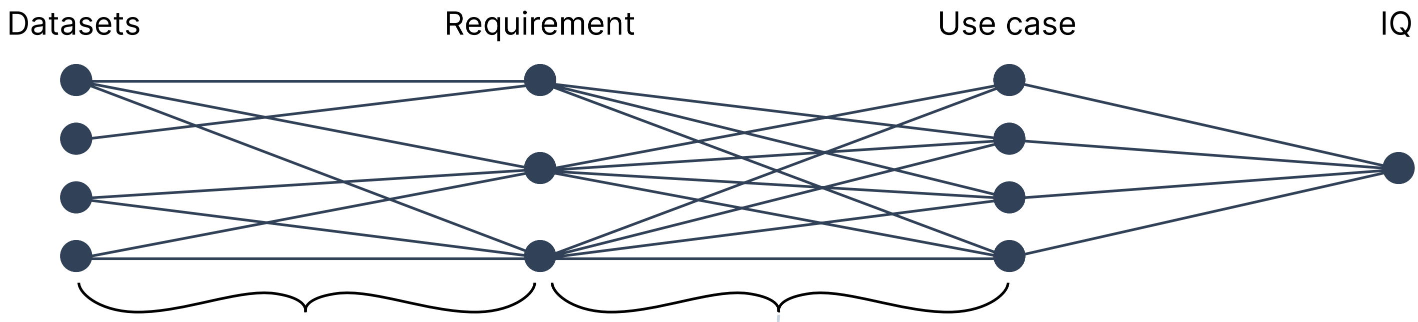
S_{GM} - *gaming*

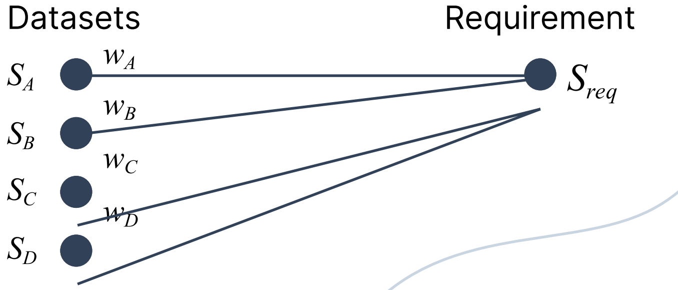
S_{OB} - *online backup*

S_{WB} - *web browsing*

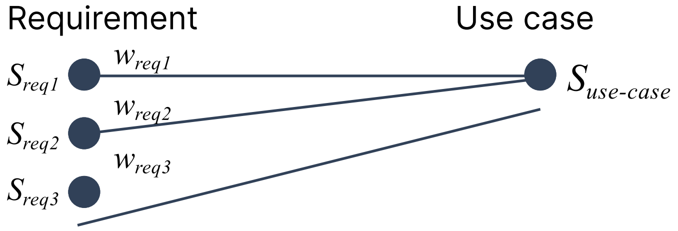
$S_{use\text{-}case}$ - Overall score for use cases

Formula cheatsheet (visual)





$$S_{req} = \frac{S_A . w_A + ... + S_D . w_D}{w_A + ... + w_D}$$



$$S_{use\text{-}case1} = \frac{S_{req1} . w_{req1} + ... + S_{req4} . w_{req4}}{w_{req1} + ... + w_{req4}}$$

Visualization Strategies

The goal of visualizing the IQB score was to show the components of what make up the score instead of just the final number. There is more to Internet quality than just the speed score which is commonly shown.

The comparison example was how credit score visualizations evolved from just being a dial with one number to a doughnut chart that shows what goes into a credit score. IQB is designed to provide a singular number created from multiple components weighted in different proportions and show this all in an easy to understand visual.

One of the challenges of visualizing this data was that the emphasis was on the gap or where the metric does not meet standards. Usually with a chart, you're showing high numbers which emphasizes a large quantity, and this project goal was the exact opposite. So how do you show the absence of data and make that gap the highlight?

When designing the IQB visualization, we looked towards using common chart types that are easy to understand and simple but use a benchmark to emphasize where the metric should be for each use case so the gap was visually highlighted. We wanted something that would be recognizable and easily understandable but work with the metric that it's intended to highlight.

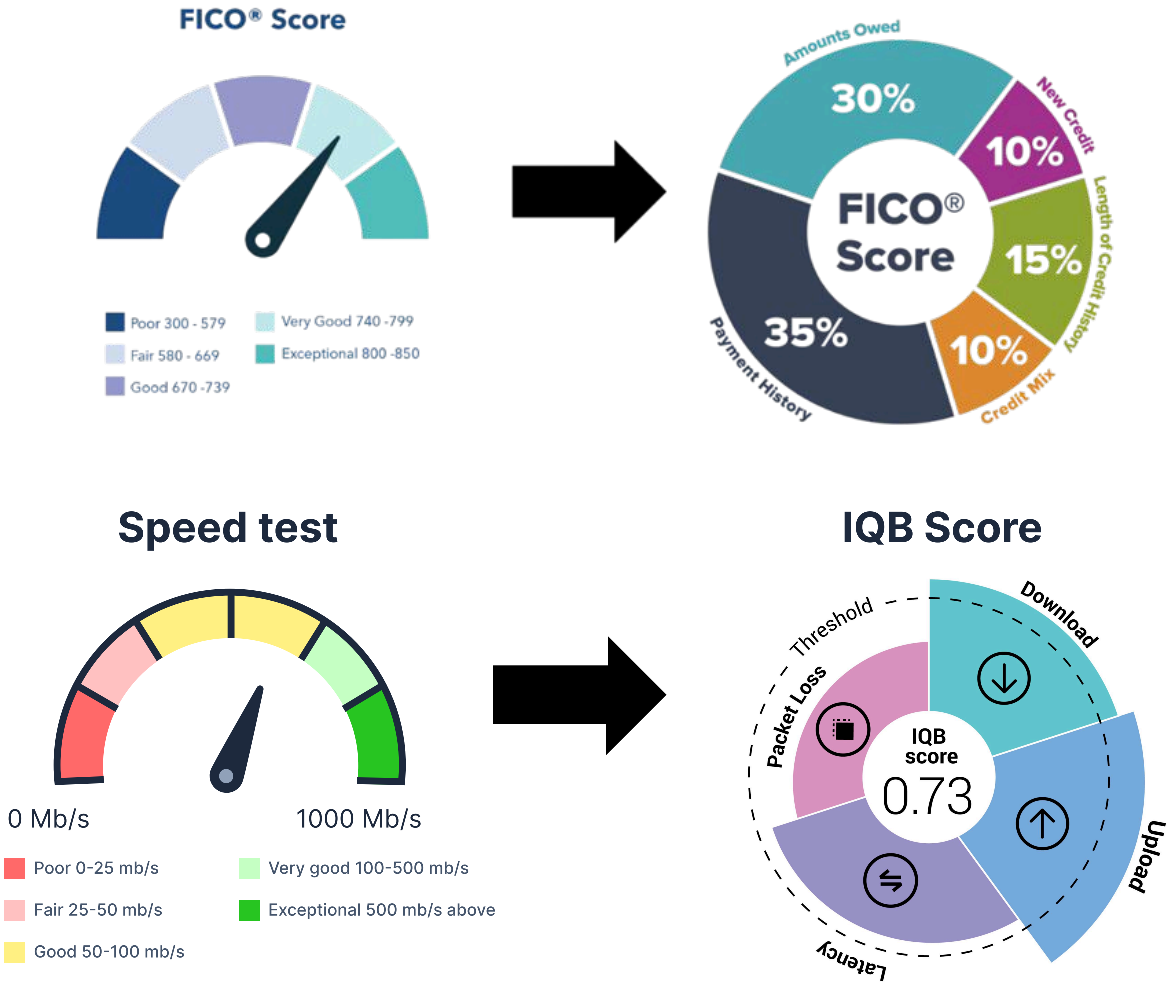


Fig 13. When building our data visualizations for IQB we considered how concepts like the credit score are visualized in an accessible way.

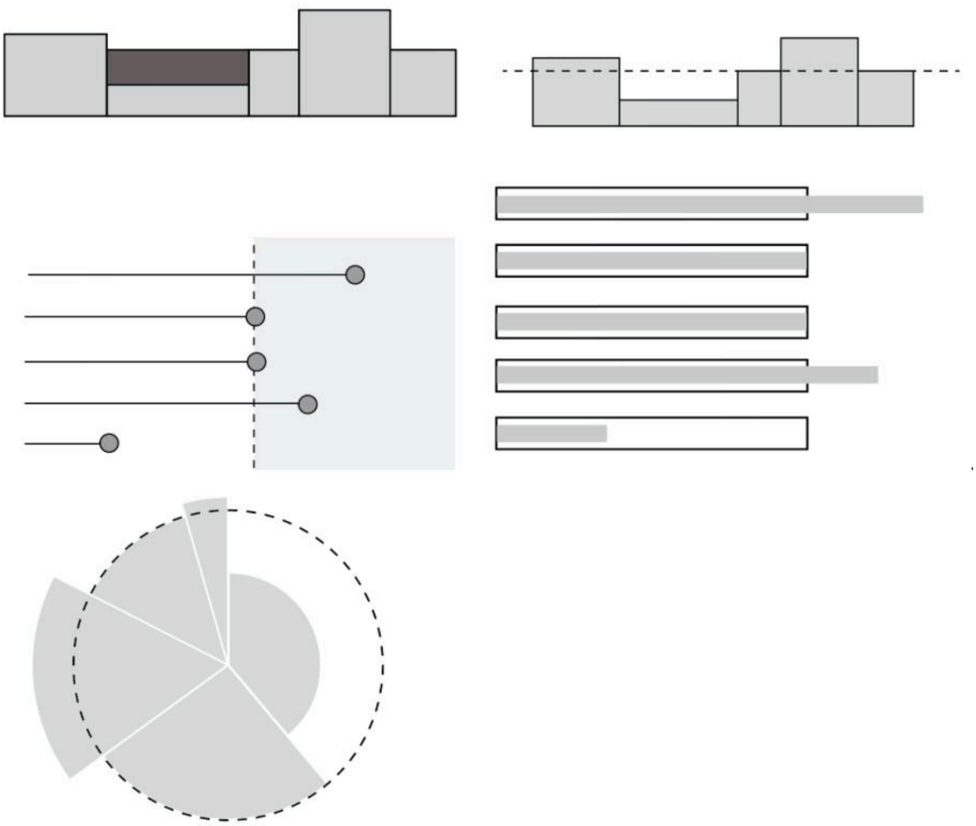


Fig 14. We experimented with several different data visualizations which would emphasize the complexity of the data, while also not overwhelming the viewer.

Ultimately we went with a combination of a variation of a sunburst chart with a benchmark line and a mosaic plot or Marimekko chart which uses both height and width encoded with data. The sunburst chart is used as the main visual and easily conveys the overall score and where the components fall short, meet or exceed the benchmark.

By showing each component that makes up the score, it shows the nuance of the metric. The target audience for this visual can see how these components fall short of the desired benchmark of effective Internet quality for the use case. A sunburst chart is fairly easy to read. It's a similar chart to a doughnut chart, but each flare is sized by the weight and extended depending on the score compared to the benchmark.

The matrix plot helps explain and give more context to the sunburst. Each component is broken down into the sources which are shown against the benchmark. Using a combination of these 2 visualizations to show the overall score and the complexity behind it brings IQB to life as an understandable metric.

Iteration

For the IQB to remain useful, it must be continuously refined. A significant limitation of many existing Internet measurement methodologies is their inability to adapt to changes in the Internet. As the Internet evolves, our measurement approaches must evolve as well, driven by a consensus-based process that prioritizes user experience, reflects the needs of our audience, and includes diverse expert stakeholders.

Differing opinions and reactions to the configuration of the IQB are not only expected but welcomed, both in its first iteration and in every iteration that follows. The IQB is designed as a flexible framework rather than a fixed standard, intentionally allowing for input, debate, and adaptation. At its core, the IQB consists of a tiered structure: use cases, network requirements, and datasets. However, the specific components within these tiers—how they are formulated into a score, the datasets used, and other foundational details—are open to discussion, feedback, and iteration.

Governance Structure and Approval Process

In order to support an iterative IQB tool that is consensus-driven, a governance structure will be created in future stages of the project. One potential structure could

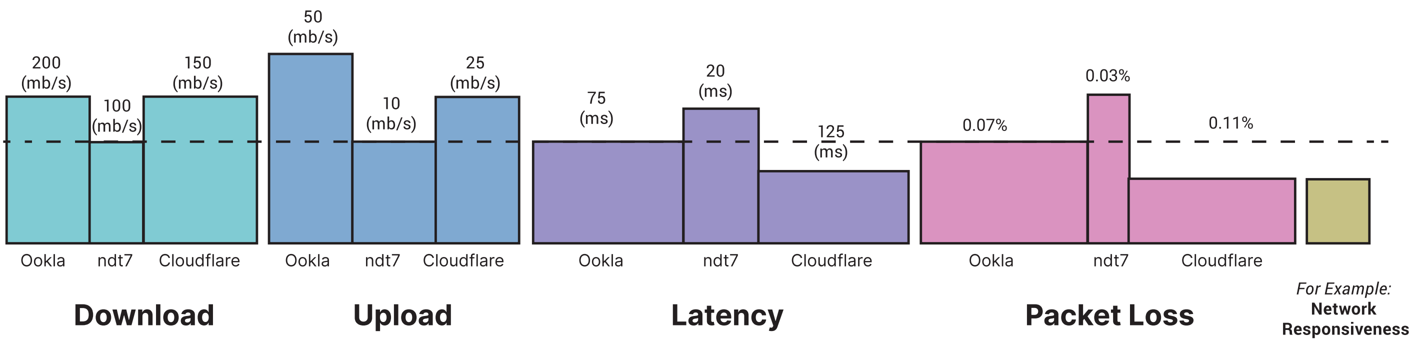


Fig 15. Contribution of datasets and network metrics to the IQB score. Bar width represents the weight of each dataset, while bar height indicates the measured value of the corresponding network requirement. The dashed line marks the minimum quality threshold for each metric.

be an IQB Advisory Committee (IQB-AC) composed of stakeholders involved in the IQB process thus far. Membership would be open to anyone, but the committee would have a responsibility to actively ensure diversity across areas of expertise, such as network measurement and policy expertise. and geographic location. For example, if there is an overrepresentation of academic network experts compared to non-academic digital inclusion advocates, efforts should be made to recruit contributors from underrepresented areas. There may even be regional subcommittees, if appropriate.

The IQB-AC would create a schedule for submission periods for proposed changes with guidance on how to submit proposals. Review periods would take place regularly for IQB-AC members to provide feedback and comments on proposals. Decisions would likely be made using digital voting mechanisms to determine consensus. IQB-AC may also have meetings to determine results for unresolved proposals.

By allowing iterations of the components, such as the formula, there will inevitably be cases where a future tool or resource, such as a map or visualization or report, that used one approved version of IQB will not match a later one. Such discrepancies can cause confusion with users. To mitigate confusion, it would be best practice for uses of IQB to always be accompanied by the version of IQB being used, e.g. IQB 2.1. Additionally, the latest IQB should always be used to reflect the most up to date version of how Internet quality is defined. Consistent updates could result in changes to the understanding of Internet quality, but in the IQB framework, this is a feature, not a bug.

Customizing IQB

As described above, IQB is designed for flexibility, meaning that the tiered structure is fixed, but all of the components within it can be iterated upon. The Iteration section discusses how proposals to IQB Global are made. This section discusses the process for forking and creating a custom version of IQB.

A custom version of IQB might be desired if the Global version of IQB prioritizes its components in a way that doesn't align with the needs of a region. For example, in a specific region, the use cases that we have proposed might not be relevant. A policymaker or advocate in that region might recognize the value in the structure of IQB, but might want to modify how the use cases are weighted in the formula such that a particular use case is weighted less or more. They might also think that their priorities are unique and do not see a need to submit an iteration for approval by the IQB-AC to change the IQB Global version.

In such a case, the regional policy maker or advocate can take the IQB formula, make the modifications they see fit and submit the version to be a verified custom version. Submissions should include information similar to that of an IQB Global submission including a description of the proposed change, its motivation, and the expected impact on IQB results. However, custom versions are not subject to the same approval process as changes to the IQB Global standard. Rather than consensus-based voting, custom versions will only be subject to a no-objections review. The goal of this lightweight process is not gatekeeper custom versions, while also encouraging shared awareness of the different versions of IQB that exist.

The Future of IQB

Proposed Next Steps

This report reflects the first stage of the IQB initiative to redefine Internet quality beyond speed. In the first stage, the primary goal was to lay a strong foundation for future iterations, tools and applications of the IQB framework that will help ensure

it can last the test of time. Prospective next stages of IQB would involve the following:

- **An IQB data collection tool.** At the moment IQB exists only as a framework for interpreting existing data collected by an unstandardized set of tools. Providing a tool for users to run tests that collects all the datasets included in IQB would greatly enhance the efficacy of the framework by making it easier for users of IQB to fill in the gaps of existing data and as a result, improve the confidence they have in its results. Existing tools such as M-Lab's Murakami and University of Chicago's Internet Equity Initiative Netrics were built to collect data with multiple methodologies and can be leveraged to collect data identified as relevant to IQB.
  - **Benchmarking and calibration.** As part of this data collection effort, IQB could be calibrated against various use cases

- **A global IQB pipeline and dashboard.** The IQB framework proposes strategies for visualization but a data pipeline and dashboard have not yet been developed to implement these strategies in a live environment. In the future, the IQB framework could enable the development of these tools and set a representative example of how others representative example of how others can create their own visualizations of the insights provided by IQB.

- **A Governance Structure.** One potential model could be an IQB Advisory Committee as described in the Iteration section. Initially the IQB-AC would be an entirely voluntary body, but over time it may require resources to maintain and administer the governance structure and make the IQB project sustainable.

Reflections and Lessons Learned

- **More global representation.** Though it was a stated interest of the project from the beginning to have a more globally inclusive representation of Internet quality, we ultimately struggled to do so through our implementation. Improvements could include including a staff member outside the US or EU and identifying partnerships with institutions in regions of interest.

- **More consistent digital communications.** Though we did multiple outreach and communications events in person, we would have liked to have had a more consistent, transparent digital communications cadence with our stakeholders and interested parties. Improvements could include a well-maintained blog, social media presence, etc.

- **More non-Internet-measurement expertise.** There are many unique challenges specific to Internet measurement, but there are also many facets to the work that deal with non-subject-matter specific details, such as data visualization and scientific communications. Given how much inspiration IQB takes from efforts like Nutri-score and the concept of credit score development, we could have facilitated more conversations and interactions with organizations building similar standards outside of the context of the Internet.

Conclusion

The Internet Quality Barometer initiative represents a critical step toward redefining how we assess Internet performance in a way that reflects the complexity of modern connectivity needs. By moving beyond the outdated reliance on speed alone, IQB introduces a more holistic framework that accounts for a range of metrics essential to real-world Internet experiences. Through extensive engagement with experts, policymakers, and industry stakeholders, this initiative seeks to create a structured, yet flexible, approach to evaluating Internet quality. The framework acknowledges the value of existing measurement tools while addressing their limitations, providing a mechanism for interpreting diverse datasets in a way that supports more informed decision-making. By embracing a tiered approach that connects user experiences with network requirements and available data sources, IQB offers a more meaningful way to gauge and improve Internet quality.

> *By moving beyond the outdated reliance on speed alone, IQB introduces a more holistic framework that accounts for a range of metrics essential to real-world Internet experiences.*

As Internet use continues to evolve, so too must the ways we measure and advocate for high-quality connectivity. The IQB framework is designed to be iterative, continuously refined based on new insights and advancements in measurement methodologies. By fostering transparency, inclusivity, and expert-driven dialogue, IQB not only enhances our understanding of Internet quality but also equips decision-makers with actionable intelligence to drive positive change. Whether through shaping policy, guiding infrastructure investments, or empowering users to advocate for better service, IQB lays the foundation for a more resilient and equitable digital future. Ultimately, by shifting the conversation away from a single dominant metric and toward a broader, more nuanced understanding of performance, IQB helps ensure that Internet quality keeps pace with the growing demands of users worldwide.

Appendix - I

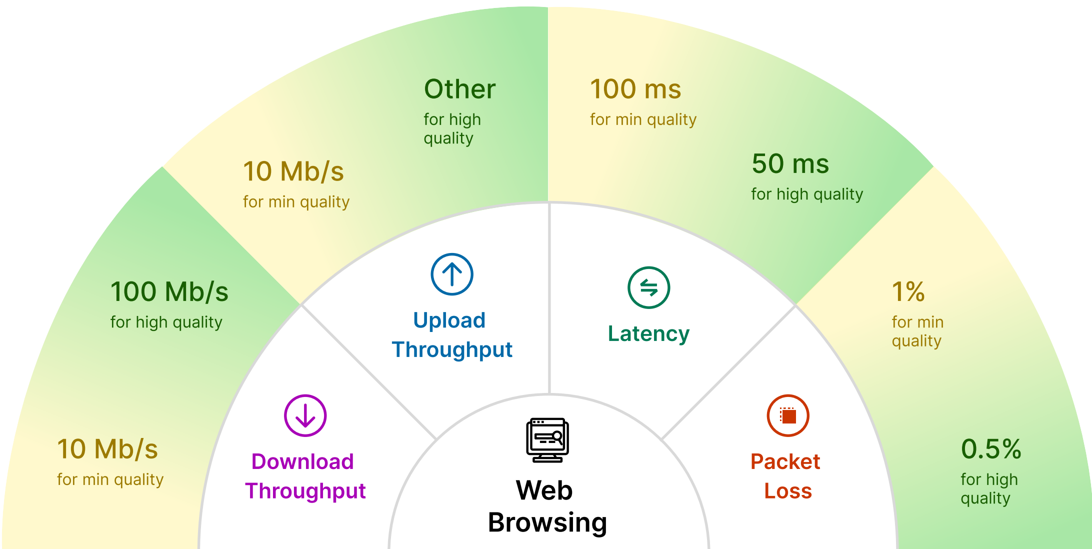
Network requirement thresholds per use case

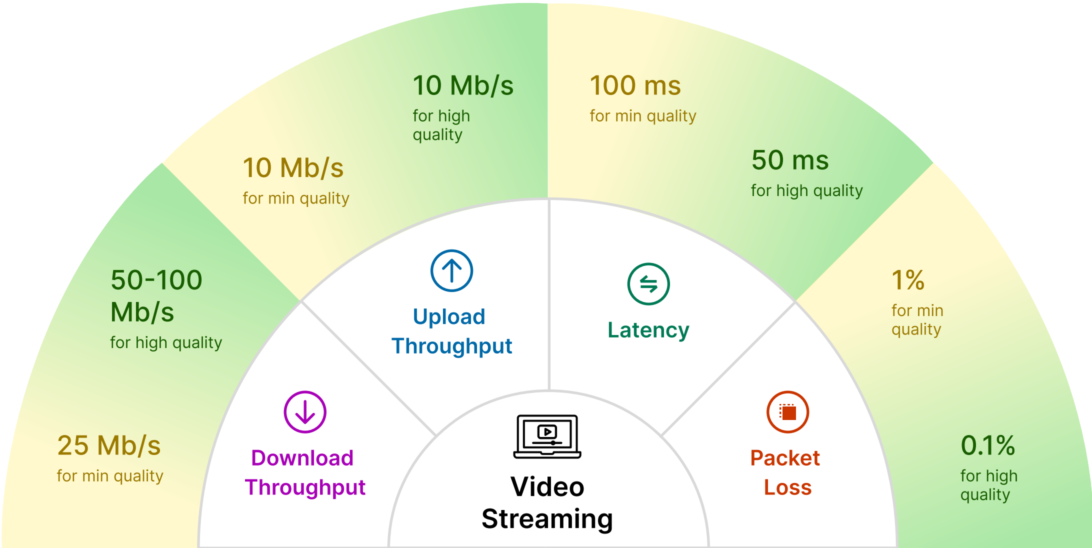
The following table presents the network requirements thresholds for minimum and high quality for each use case.

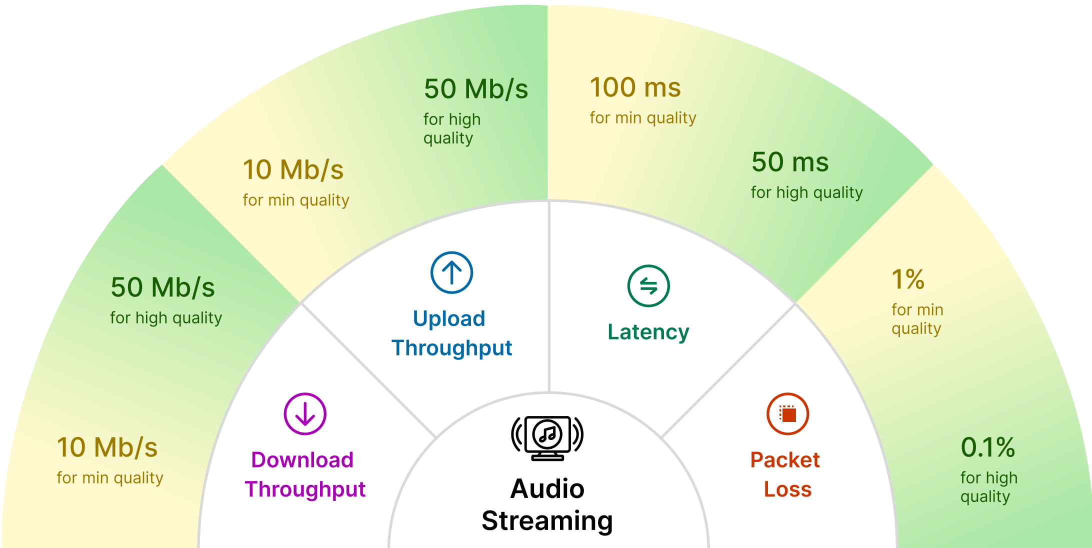
Use case	Metric	Threshold for minimum quality	Threshold for High quality
Web Browsing	Download Throughput	10 Mb/s	100 Mb/s
	Upload Throughput	10 Mb/s	other
	Latency	100 ms	50 ms
	Packet Loss	1%	0.5%
Video Streaming	Download Throughput	25 Mb/s	Tie between 100 and 50 Mb/s
	Upload Throughput	10 Mb/s	10 Mb/s
	Latency	100 ms	20 ms
	Packet Loss	1%	0.1%
Audio Streaming	Download Throughput	10 Mb/s	50 Mb/s
	Upload Throughput	5 Mb/s	10 Mb/s
	Latency	100 ms	50 ms
	Packet Loss	1%	0.1%
Video Conferencing	Download Throughput	25 Mb/s	100 Mb/s
	Upload Throughput	25 Mb/s	100 Mb/s
	Latency	50 ms	20 ms
	Packet Loss	0.5%	0.1%

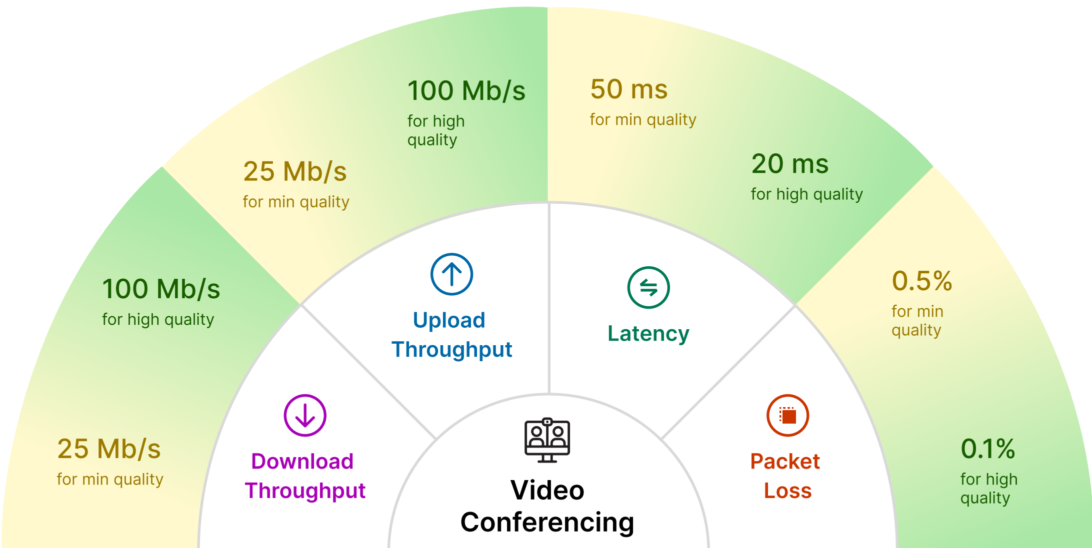
Online Backup	Download Throughput	10 Mb/s	10 Mb/s
	Upload Throughput	25 Mb/s	200 Mb/s
	Latency	100 ms	100 ms
	Packet Loss	1%	0.1%
Gaming	Download Throughput	25 Mb/s	100 Mb/s
	Upload Throughput	25 Mb/s	100 Mb/s
	Latency	10 ms	1 ms
	Packet Loss	0.5%	0.1%

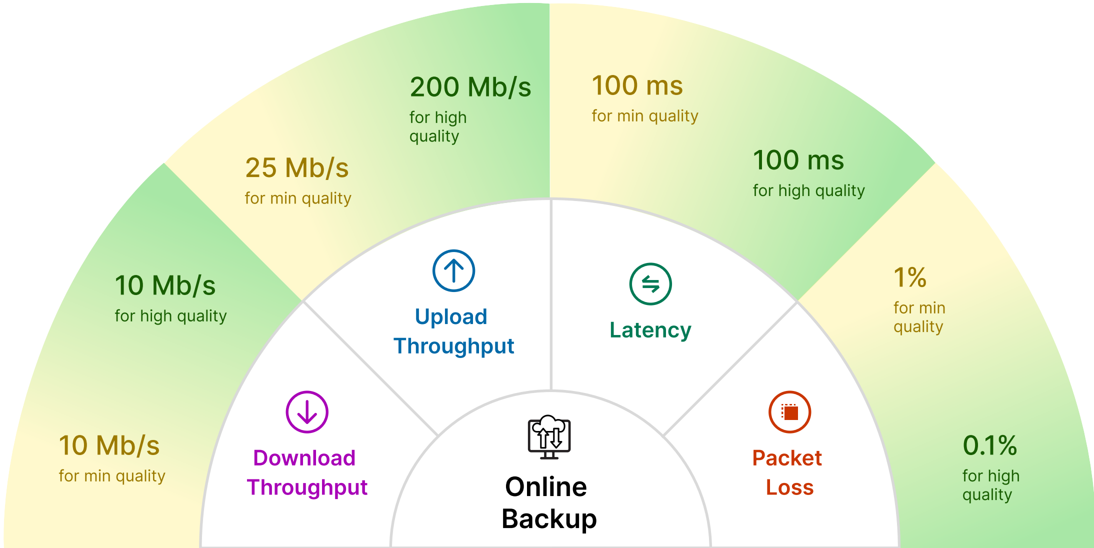
The following figures depict the network requirements thresholds for minimum and high quality (i.e., the same information as in the table above) for each use case individually.



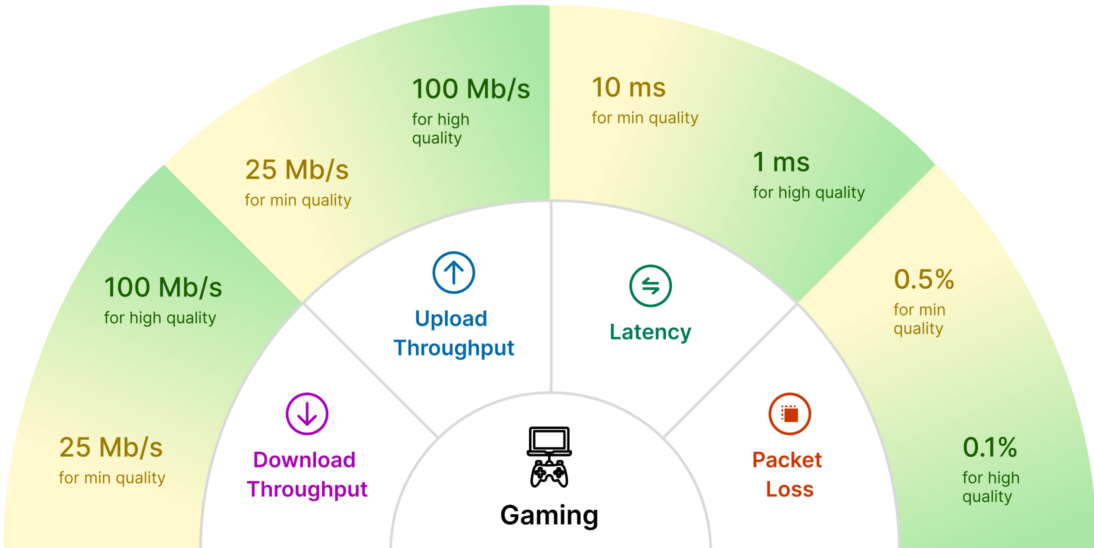

Video Streaming
Download Throughput
25 Mb/s for min quality
50-100 Mb/s for high quality
Upload Throughput
10 Mb/s for min quality
10 Mb/s for high quality
Latency
100 ms for min quality
50 ms for high quality
Packet Loss
1% for min quality
0.1% for high quality


Audio Streaming
Download Throughput
10 Mb/s for min quality
50 Mb/s for high quality
Upload Throughput
10 Mb/s for min quality
50 Mb/s for high quality
Latency
100 ms for min quality
50 ms for high quality
Packet Loss
1% for min quality
0.1% for high quality


Video Conferencing
Download Throughput
25 Mb/s for min quality
100 Mb/s for high quality
Upload Throughput
25 Mb/s for min quality
100 Mb/s for high quality
Latency
50 ms for min quality
20 ms for high quality
Packet Loss
0.5% for min quality
0.1% for high quality



200 Mb/s
for high quality
25 Mb/s
for min quality
10 Mb/s
for high quality
10 Mb/s
for min quality
100 ms
for min quality
100 ms
for high quality
1%
for min quality
0.1%
for high quality
Upload Throughput
Latency
Download Throughput
Packet Loss
Online Backup



100 Mb/s
for high quality
25 Mb/s
for min quality
100 Mb/s
for high quality
25 Mb/s
for min quality
10 ms
for min quality
1 ms
for high quality
0.5%
for min quality
0.1%
for high quality
Upload Throughput
Latency
Download Throughput
Packet Loss
Gaming

Appendix - II

Network requirement weights per use case

User case	Metric	Weight (1 to 5)
Web Browsing	Download speed	3
	Upload speed	2
	Latency	4
	Packet loss	4
Video Streaming	Download speed	4
	Upload speed	2
	Latency	4
	Packet loss	4
Audio Streaming	Download speed	4
	Upload speed	1
	Latency	3
	Packet loss	4
Video Conferencing	Download speed	4
	Upload speed	4
	Latency	4
	Packet loss	4
Online Backup	Download speed	4
	Upload speed	4
	Latency	2
	Packet loss	4
Gaming	Download speed	4
	Upload speed	4
	Latency	5
	Packet loss	4

Appendix - III

Detailed Description of Formula

To build an IQB score, we can start at the "bottom" of the layers with the Datasets tier, and work our way upwards. In this example, we'll use a high quality threshold, though the same process could be applied to the minimum quality thresholds as well.

Datasets

Dataset Aggregations

Assume that we have four datasets with multiple test results in the region or network of interest. Datasets A, B, and C return "classic" metrics e.g. download, upload, latency and packet loss. Dataset D returns an "advanced summary" metric e.g. "Low, Medium or High".

Let's also assume that each dataset has a statistically significant sample size e.g. 1000 tests.

First, we take the 95th percentile of each dataset. For datasets with classic metrics, this means we take the 95th percentile for download, upload, latency and packet loss, and for advanced summary metrics we take the 95th percentile of the results in their chosen format e.g. the 95th percentile of clients in the sample of Dataset D received a rating of "High".

For example, the datasets could produce the following aggregations (see table).

Dataset Source	95th Percentile Download	95th Percentile Upload	95th Percentile Latency	95th Percentile Packet Loss	95th Percentile Advanced Summary Metric
A	>= 100 Mb/s	>= 10 Mb/s	<= 20 ms	<= 0.1%	-
B	>= 100 Mb/s	>= 10 Mb/s	<= 20 ms	<= 0.1%	-
C	>= 100 Mb/s	>= 10 Mb/s	<= 20 ms	<= 0.1%	-
D	-	-	-	-	High

Binary Requirement Score

The aggregations from each of these datasets can be converted into a binary requirement value that indicates whether or not the aggregation from a given dataset indicates that the threshold for that individual network requirement for a high-quality experience of a particular use case is met.

The binary requirement value can be TRUE or FALSE. A value of TRUE indicates that the quality threshold was met or exceeded. A value of FALSE indicates that it did not meet the threshold.

Network Requirement Thresholds

As described above, the following thresholds have been defined for a high-quality experience for the video streaming use case. If a test result meets or exceeds the threshold, its binary requirement value will equal TRUE.

Combining the information from the two tables above (Dataset aggregations and Network Requirement Thresholds) Datasets A-D would produce the following binary requirement scores for the Video Streaming use case.

can be reduced to ensure it impacts the overall score less.

For datasets that provide classic metrics, the requirement agreement score is an aggregation of each dataset's binary requirement value and its associated weight.

For test results that provide summary metrics, there is no need for aggregation because the metric is a unique value. The requirement agreement score is therefore equivalent to its binary requirement value.

Dataset Source	Binary Requirement Scores				
	Download	Upload	Latency	Packet Loss	Advanced Summary Metric
A	TRUE	TRUE	TRUE	FALSE	-
B	TRUE	TRUE	FALSE	FALSE	-
C	FALSE	FALSE	TRUE	TRUE	-
D	-	-	-	-	TRUE

Requirement Agreement Score

The binary requirement values can be combined into a **requirement agreement score**, which indicates the degree to which the available datasets "agree" that a network requirement for a high-quality user experience for a given use case is being met.

Each dataset is also given a weight, which indicates how relevant that dataset is to that network requirement for that particular use case, and allows for certain datasets to impact the overall score more or less. For example, if a dataset has a known limitation for its measurement of packet loss for all use cases, or one in particular, its weight

Let's assume we've defined the following weights for each dataset for latency while video streaming can be reduced to ensure it impacts the overall score less (refer to the figure on the next page).

For datasets that provide classic metrics, the requirement agreement score is an aggregation of each dataset's binary requirement value and its associated weight.

For test results that provide summary metrics, there is no need for aggregation because the metric is a unique value. The requirement agreement score is therefore equivalent to its binary requirement value.

Example: Dataset Weights - Latency While Video Streaming

Let's assume we've defined the following weights for each dataset for latency while video streaming

Weight	0.25	0.25	0.5
Dataset	A	B	C

Requirement Agreement Score Formula

The per-requirement agreement score for each network requirement uses the following formula:

$$S_{req} = \frac{(w_A \times S_A) + (w_B \times S_B) + \ldots}{w_A + w_B + \ldots} = \frac{(w_A \times S_A) + (w_B \times S_B) + \ldots}{1}$$

Where:

- S_{req} is the overall agreement score for the network requirement e.g. latency or packet loss.
- w_A w_B ... are the weights of the individual datasets for the network requirement.
- S_A S_B ... are the binary requirement values per dataset for the network requirement.

A rule of thumb for evaluating the Requirement Agreement Score can be:

≥ 0.7 — A requirement agreement score ≥ 0 .7 indicates that there is sufficient evidence within the available datasets that the network requirement for the given use case is being met.

< 0.7 — A requirement agreement score < 0.7 indicates that there is not sufficient evidence within the available datasets that the network requirement for the given use case is being met.

Example: Binary Agreement Score - Datasets A-C - Latency for Video Streaming

For example, to evaluate how strongly the available datasets agree that the network requirements for latency is being met for high-quality experience video streaming, and using the values in the tables above (weights: 0.25, 0.25, and 0.5, and binary requirement scores TRUE, FALSE, TRUE for datasets A,B, and C, respectively), we get:

$$S_{latency} = \frac{(0.25 \times 1) + (0.25 \times 0) + (0.5 \times 1)}{0.25 + 0.25 + 0.5}$$
$$S_{latency} = 0.75$$

- From this result, we could say that there **is** sufficient evidence within the available datasets that the network requirements for latency are being met for a high-quality video streaming experience.

Example: Binary Agreement Score - Datasets A-C - Packet Loss for Video Streaming:

For another example, let's assume the same weights (0.25, 0.25, and 0.5) for each dataset for packet lossand the binary requirement scores FALSE, FALSE, and TRUEwhile video streaming. Then we get:

$$S_{packet-loss} = \frac{(0.25 \times 0) + (0.25 \times 0) + (0.5 \times 1)}{0.25 + 0.25 + 0.5}$$
$$S_{packet-loss} = 0.5$$

- From this result, we would say that there **is not** sufficient evidence within the available datasets that the network requirements for packet loss are being met for a high-quality video streaming experience.

Example: Binary Agreement Score - Dataset D for Video Streaming

For our example using Dataset D for the evaluation of high-quality video streaming, we use the binary requirement value directly.

$$S_{D} = 1.0$$

Use cases

Use case score

These binary requirement values can then be aggregated to produce a use-case score, which indicates the degree to which the available datasets indicate that users in the region or network will have a high-quality experience of a use case.

Each network requirement is also given a weight, which tells us how important that network requirement is to that use case and allows for certain network requirements to impact the overall score more or less. For example, if a given dataset is designed to measure a particular use case, it could be given a higher weight.

The per use case score uses the following formula:

$$S_{use-case} = \frac{(w_{req1} \times S_{req1}) + (w_{req2} \times S_{req2}) + \ldots}{w_{req1} + w_{req2 +} + \ldots}$$

Where:

- $S_{use-case}$ is the overall score for the use case.
- w_{req1} w_{req2} are the weights of the individual network requirements for the use case.
- S_{req1} S_{req2} ... are the weighted scores for the network requirements.

To evaluate the download, upload, latency and packet loss metrics produced by Dataset A-C and Dataset D's advanced summary metric as the network requirements, the formula is as follows:

$$S_{use-case} = \frac{(w_{DL} \times S_{DL}) + (w_{UL} \times S_{UL}) + (w_{L} \times S_{L}) + (w_{PL} \times S_{PL}) + (w_{D} \times S_{D})}{w_{DL} + w_{UL} + w_{L} + w_{PL} + w_{D}}$$

Use Case Score Evaluation

This resulting single score for each use-case is calculated using a weighted aggregation of the network requirements scores.

≥ 0.7 — Indicates a sufficient amount of the network requirements were met for a given use case, and therefore the available data does indicate a high-quality experience in the network or region for the given use case.

< 0.7 — This does not indicate sufficient amount of the network requirements were met for a given use case, and therefore the available data does not indicate a high-quality experience in the network or region for the given use case.

Example: Use Case Score - Video Streaming

In the above examples we determined the following requirement agreement scores:

- $S_{L} = 0.75$
- $S_{PL} = 0.5$

- $S_D = 1.0$

Let's also assume that we've found the following agreement scores for download and upload:

- $S_{DL} = 0.75$
- $S_{UL} = 0.5$

Let's assume we've defined the following weights for each network requirement for a given use-case, e.g. video streaming:

Network Requirement	Formula Value	Weight
Download (DL)	w_{DL}	4
Upload (UL)	w_{UL}	2
Latency (L)	w_L	4
Packet Loss (PL)	w_{PL}	4
Dataset D	w_D	4

The calculation of the use-case score for video streaming is then as follows:

$$S_{video-streaming} = \frac{(4 \times 0.75) + (2 \times 0.5) + (4 \times 0.75) + (4 \times 0.5) + (4 \times 1.0)}{4+2+4+4+4} = 0.72$$

▶ With these results, we can say the available data **does** indicate a high-quality experience in the network or region for the video-streaming use case.

Internet Quality

Internet Quality Score

Each test result will produce six different use-case scores, one for each use case selected by IQB:

- Video Streaming: VS
- Audio Streaming: AS
- Video Conferencing: VC
- Gaming: GM
- Online Backup: OB
- Web Browsing: WB

The Internet Quality Score combines the use case scores provided by each test result. Each use case is also assigned a weight ($w_{use\text{-}case}$), which indicates how significant it is to Internet quality.

The formula for the Internet Quality Score for an individual test is as follows:

$$S_{IQ} = \frac{(w_{VS} \times S_{VS}) + (w_{AS} \times S_{AS}) + (w_{VC} \times S_{VC}) + (w_{GM} \times S_{GM}) + (w_{OB} \times S_{OB}) + (w_{WB} \times S_{WB})}{w_{VS} + w_{AS} + w_{VC} + w_{GM} + w_{OB} + w_{WB}}$$

Evaluating the Internet Quality Score

≥ 0.7

An Internet Quality Score ≥ 0.7 indicates a sufficient amount of the use cases were indicated to be high-quality, and therefore the available data does indicate a high-quality Internet experience for the region or network of interest.

< 0.7

An Internet Quality Score < 0.7 does not indicate a sufficient amount of the use cases were indicated to be high-quality, and therefore the available data does not indicate a high-quality Internet experience for the region or network of interest.

Use Case Weights

Let's assume we've given each use case an equal weight, i.e., ⅙.

Example: Datasets A-D - Internet Quality Score

Let's assume we've calculated the use–case score for the remainder of the use cases, resulting in the following:

- $S_{AS} = 0.8$
- $S_{VC} = 0.8$
- $S_{GM} = 0.8$
- $S_{OB} = 0.6$
- $S_{WB} = 0.6$

Example: Datasets A-D - Internet Quality Score

As shown above, S_{VS} = 0.72

$$S_{IQ} = \frac{(\frac{1}{6} \times 0.72) + (\frac{1}{6} \times 0.8) + (\frac{1}{6} \times 0.8) + (\frac{1}{6} \times 0.8) + (\frac{1}{6} \times 0.6) + (\frac{1}{6} \times 0.6)}{\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}} = 0.72$$

With these results we can conclude, the available data from Datasets A-D does indicate a high-quality Internet experience for the region or network of interest.

Acknowledgements

We would like to acknowledge the team behind this project: Lai Yi Ohlsen led the research and development of the project; Melissa Newcomb provided program management and grant administration; Pavlos Sermpezis provided analytical and technical input; John Horrigan supported formula development; Amy Cesal designed technical visual graphics, Abishek Sharma led the publication and visual design. The IQB research project would not be possible without the input of diverse stakeholders who participated throughout the project. The following people contributed to the IQB project by participating in one or more of the following: key stakeholder interview(s), workshop(s), online survey, or provided feedback on the project.

Amreesh Phokeer, Internet Society
Andra Lutu, Senior Researcher, Telefonica
Anna Brunstrom, Karlstad University
Arpit Gupta, Assistant Professor, UCSB
Bjørn Ivar Teigen Monclair, Head of Research, Domos
Brandon Schlinker, Research Scientist, Meta
David Belson, Cloudflare
David Clark, Senior Research Scientist, Massachusetts Institute of Technology
Dr. Nick Merrill, Director, Daylight Lab, UC Berkeley Center for Long-Term Cybersecurity
Edmundo de Souza e Silva, Professor, Systems Engineering and Computer Science/ COPPE, Federal University of Rio de Janeiro
Elizabeth Belding, Professor, UC Santa Barbara
Emile Aben, RIPE NCC
Ethan Katz-Bassett, Associate Professor, Columbia University
Georgia Bullen, Superbloom Design & Measurement Lab
Juan Peirano, Internet Society
Keith Winstein, Stanford University
Kevin Vermeulen, LIX, CNRS, Ecole Polytechnique
Lauren Crean, OECD
Le-Tian Cheng, Georgia Institute of Technology
Loqman Salamatian, Columbia University
Luis M. Contreras, Telefónica, CTIO Unit
Luke Stringer, Giga (UNICEF-ITU)

Marcos Schwarz, R&D Manager RNP - Brazilian National Research and Education Network
Maria Antonia Bravo, Tech Programme Manager, Giga (UNICEF-ITU)
Marwan Fayed, Cloudflare and University of St Andrews
Matt Calder, Engineer, Meta
Mike Klemencic, Principal Engineer, N3 Laboratories
National Association of Counties (NACo)
Neil Davies, Predictable Network Solutions Ltd
Nick Feamster, University of Chicago
Nick Merrill, Research Scientist, UC Berkeley Center for Long-Term Cybersecurity
Peter Boothe, Measurement Lab
Prof. Vijay Sivaraman, UNSW Sydney
Raza Panjwani, Senior Policy Counsel, Open Technology Institute at New America
Revati Prasad, Vice President of Programs, Benton Institute for Broadband & Society
Ricky Mok, CAIDA/UC San Diego
Robert Kisteleki, Principal Engineer, RIPE
Roberto D'Auria, Platform Engineer, Measurement Lab
Romain Fontugne, Deputy Directory, IIJ Research Laboratory
Samantha Schartman, Director of Philanthropic Programs, Connect Humanity,
Sascha Meinrath, Director, X-Lab, Co-Founder, M-Lab, Palmer Chair in Telecommunications, Penn State University
Scott Jordan, Professor, University of California Irvine
Shaddi Hasan, Assistant professor, Virginia Tech
Simon Sundberg, Karlstad University
Simone Ferlin-Reiter, Red Hat, KarlstadU
Soledad Luca de Tena, Consultant, Giga (UNICEF-ITU)
Syed Tauhidun Nabi, Virginia Tech
Timur Friedman, Sorbonne University
Varshika Srinivasavaradhan, UC Santa Barbara
Vasanta Chaganti, Swarthmore College
Vasileios Giotsas, Research Engineer, Cloudflare
Warren Kumari, Google
Zachary Bischof, Georgia Tech