Measuring Mobile Broadband
Challenges and Implications for Policymaking

Zoraida Frias[1]
Universidad Politécnica de Madrid

William Lehr
MIT & Weizenbaum Institute[2]

Volker Stocker[3]
Weizenbaum Institute & TU Berlin

Luis Mendo
Universidad Politécnica de Madrid

Version: August 2023

Abstract

Mobile broadband networks constitute essential infrastructure to enable a wide range of innovative services and use cases anticipated for our digital economy future. Measuring performance is essential in many ways. *First*, to allow service providers to manage and develop their networks. *Second*, for the efficient operation of markets, and *third*, for evidence-based policymaking. In the rapidly evolving digital economy, capabilities for collecting more fine-grained measurements and analytics that deliver insights to enable real-time network management and localized control are expanding. As the fundamental methods used to collect measurement data are changing, the ecosystem of stakeholders with strategic interests in mobile measurement is growing and becoming more complex, posing challenges and opportunities for policymakers. Against the background of this growing complexity, this paper aims to discuss some basic features of a capable and reliable measurement ecosystem for mobile broadband. We document how the mobile broadband measurement ecosystem has changed and discuss its implications on a number of important broadband policy issues.

Keywords
Mobile Broadband, Measurement, Performance, Evidence-based policymaking, Internet Policy, Telecommunications Policy

---

[1] Corresponding author: Avda. Complutense 30, 28040 Madrid, zoraida.frias@upm.es. Zoraida Frias would like to express her gratitude to the European Commission for partial funding through the H2020 project ARIADNE under Grant Agreement 871464.

[2] William Lehr is an associated researcher in the research group "Digital economy, Internet ecosystem, and Internet Policy" at the Weizenbaum Institute for the Networked Society.

[3] Volker Stocker would like to acknowledge funding by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII131 (Weizenbaum-Institut für die vernetzte Gesellschaft—Das Deutsche Internet-Institut).

Table of Contents

1. Introduction ... 3
2. Changing Measurement Ecosystem ... 7
   2.1. The growing complexity of measuring mobile broadband performance ... 8
   2.2. Performance Measurements for Mobile Broadband Services ... 11
   2.3. Evolution of Drive/Walk Tests ... 14
   2.4. Evolution of Mobile Crowdsourced Measurements ... 17
   2.5. The State and Future Trajectory of the Measurement Ecosystem — Reflections and Discussion ... 21
3. Discussion & Implications for Policymaking ... 23
   3.1. Universal service and connectivity targets ... 26
   3.2. Consumer information asymmetries ... 30
   3.3. Other Challenges: Value Chain and Resource Sharing Dynamics and Concerns Regarding Privacy and Security ... 33
4. Conclusion & Outlook ... 36
References ... 41
Appendix ... 47

1. Introduction

The evolution of mobile broadband and the Internet ecosystem gives rise to novel challenges regarding measurements and their implications for evidence-based policymaking. Today, a much greater range of measurements is available to industry participants, consumers, and policymakers interested in assessing various dimensions of economic and technical performance.[4] As advancements in measurement capabilities have significantly broadened the range of data that can be collected during normal operations (or could be collected in response to specific queries), evidence that previously was simply not available can now be accessed. This, in turn, opens new possibilities for gathering valuable information for multiple stakeholders. At the same time, the mobile broadband ecosystem has grown more complex. An expanded set of industry participants offers a wider range of services to a wider array of customers. Whereas many more actors have a vested interest in measurement data to inform their decision-making, identifying the right measurements to answer particular questions is increasingly challenging and context-specific. Moreover, the potential that faulty measurements, metrics, or interpretations may result in bad decisions poses a significant risk for sound policymaking.

To better understand how the measurement landscape for mobile broadband has changed, it is worthwhile considering the needs motivating the collection of more detailed measurement data by all the ecosystem stakeholders. Moreover, data collected for one purpose (e.g., in the normal course of managing a mobile network in real time) may be used for another purpose (e.g., to compare the performance of different service providers or to inform investment planning or design regulatory interventions). Collecting, analyzing, and sharing measurement data is costly, and who gets access to what information can have strategic market and policy relevant implications (e.g., related to privacy, security, and other policy issues). In the following, we highlight three decision-making contexts that motivate the need for collecting more detailed performance data.

First, network service providers need technical performance measurements to monitor and manage their networks. The more complex the system, the more detailed measurements are needed to operate it; and the more real-time control to customize network operations is desired, the more real-time measurements are needed. Those measurements are necessary for evaluating whether providers are meeting their contractual obligations to business partners and their customers (e.g., are they meeting their Service Level Agreements (SLAs)?). They

[4] 'Performance' may be interpreted as an over-arching concept related to how a system compares relative to a standard or question, and that might be from multiple perspectives. The 'performance' of mobile broadband can mean many things in different contexts. It could be a question about the economic performance of markets (e.g., are they adequately competitive? Are consumers getting what they expect? Is quality improving over time?), or a question about technical performance, which might also refer to a wide range of questions (e.g., how is the service operating with respect to specific metrics such as throughput, latency, or reliability?). Herein, we will try to be clear whether we are discussing technical or economic performance metrics, but when used by itself, we will be referring mostly to technical operating characteristics, i.e., how do technical measurements of a provisioned service compare with some standard, while recognizing that measurements of operating performance may inform other questions like the availability of broadband service or whether traffic management at interconnection points is 'reasonable' (according to some regulatory standard).

are also needed for network planning and optimization, including regular network maintenance, fault diagnosis and recovery operations, and planning future capacity expansions. Additionally, those technical measurements are coupled to economic metrics to make sure the network operations and investments meet budget targets and are optimized to minimize costs. Mobile Network Operators (MNOs) are continuously making decisions about what activities to undertake in-house versus to outsource to third-party firms (i.e., make-vs-buy decisions). That is true about all of a network operator's network-related and non-network related business functions (from network operations to human resource management). It is also true about a network operator's measurement and non-measurement-related operations. Increasingly, the design and provisioning of the technical performance measurement capabilities is, itself, a business function that network operators need to manage. Even in a world without any competition or complex industry value chains, such performance measurements would be necessary.[5] Moreover, as the industry structure, value chains, and the range of products and services offered get more diverse, the need for more complex performance measurements increases.

Second, markets need performance measurements. Such measurements are essential for the upstream participants in the industry value chain and the retail customers (who provide the final demand driving the value chain). Firms and consumers rely on these measurements and insights to make informed decisions about what assets to invest in, what services to purchase, and how to use those services. In addition to technical performance benchmarking measurements to compare the services offered by different providers over time and across geographic or customer-segment markets, there is also a need for economic performance metrics such as market share, pricing trends, and customer satisfaction or awareness surveys. In the retail market, some customers get better service than others, where 'better' needs to be measured with appropriate attention to price-quality trade-offs and consumer choice. Mobile broadband service providers offer multiple tiers of service contracts that differ both with respect to technical performance (e.g., peak data rates, roaming capability, traffic management) and economic performance (e.g., pricing and how services are bundled with other ancillary services such as device subsidies, cloud storage, or applications). A healthy market ought to provide consumers with choices and different customers may self-select offerings with different technical and economic performance characteristics. Consumers need to know how to compare the price-quality features of different broadband and Internet offerings to decide which to subscribe to and how to use their subscriptions. Industry participants need performance measurements to make informed decisions when contracting for services at different points of the value chain and to inform their strategic decision-making regarding market entry, product marketing (and pricing), how to respond to competitors, etc.

---

[5] Industry value chains separate the production of goods and services into stages from raw materials to final goods. The Internet ecosystem value chain often differentiates products and services into business-to-business (B2B) intermediate goods and business-to-consumer (B2C) final goods. Whether there is competition or not, consumers need performance measurements to evaluate their choice of service to purchase and to monitor their satisfaction with their decision and to moderate their behavior. Upstream providers of intermediate goods and services, and complementary products and services need performance measurements to evaluate market conditions and to evaluate their buy/sell decisions. The performance measurements and concerns of different actors and their sophistication and options for evaluating those measurements differ across the value chain and at each stage, across market participants.

Third, policymakers need performance measurements to engage in sound evidence-based policymaking. They need such evidence whether they are seeking to proactively craft effective, market-based regulations and ensure compliance with existing policies, or reactively in their role adjudicating disputes among industry participants. Performance measurement data are needed to appropriately diagnose problems and assess liability when problems have been identified. Policymakers may use measurement information to assess coverage or quality gaps in what is currently being delivered and what policymakers believe should be delivered (e.g., is the quality of mobile broadband sufficiently high and on an improvement trajectory that meets national goals?). Broadband measurements are also needed to ensure that MNOs are adhering to specific performance and licensing obligations, including their spectrum licenses (e.g., are MNOs meeting build-out coverage and service quality requirements?).

All three of the motivations for collecting and making use of performance data articulated above are inter-related and co-dependent. At a fundamental level, this is a direct result of the overarching goal in the majority of economic regulatory policymaking to enable and promote efficient, market-based regulation instead of top-down government control. For communications policy, the goal is to promote policies that will encourage healthy investment and competition among digital infrastructure and service providers that will keep pace with the expanding needs of the digital economy.

The need for performance measurements for telecommunication services is not new. Before there was mobile broadband, there were fixed broadband services, and before that there were fixed-line telephony and cable television services. Measurements were always necessary, but measurement needs have become more complex over time. In earlier, simpler times, the range of services to be measured and assessed for availability, competitive effects, or technical efficiency were limited. Telephone companies offered telephony and cable television companies offered multiple channels of one-way video programming. With the transition to fixed broadband platforms, options for intermodal competition expanded and the range of services offered by broadband providers operating technically-different networks (e.g., DSL-based versus Cable modem-based access networks) expanded, necessitating a richer set of measurements to assess the performance of the bundles of services offered by different providers, the largest of which offered triple plays of fixed telephony, video programming, and broadband access services to their mass market customers. With the rise of mobile broadband, the measurement challenges became even more complex. Comparing the performance of mobile services (its availability, Quality of Service (QoS), and other service dimensions) is inherently more complicated than assessing the performance of fixed-location services. Hence, the measurement challenges, capabilities, and choices feasible for measurements to be used in evaluating the performance of mobile broadband services is even more complex than for earlier generations of telecommunication networks.[6]

---

[6] As we explained above, value chains and the range of products and services offered get more complex and diverse, thus creating an increased need for more complex performance measurements. On the one hand, the set of questions that need to be answered with the support of appropriate measurements has broadened. This, in turn, points to the need for different measurement strategies and data insights, which ultimately require a diverse and flexible measurement ecosystem. On the other hand, combining and making sense of complementary measurement data collected by different entities using different methods to gain meaningful insights is not trivial.

In previous works, we have written about the challenges of measuring the performance of fixed broadband services and the significant challenges that such measurements pose, whether one is focusing on evaluating the data rates (speed) of broadband, or something much harder, the state of competition in broadband services, or the reliability of broadband.[7] All of those same problems and challenges also arise with mobile broadband (and we will note those in the following where appropriate). In short, compared to fixed broadband services, a focus on mobile services further complicates the measurement challenge due to mobility and the nature of the wireless channel.[8]

Against the background of this growing complexity, this paper aims to discuss some basic features of a capable and reliable measurement ecosystem for mobile broadband. First, we document how the mobile broadband measurement ecosystem has evolved and discuss the challenges of measuring mobile network performance considering new technologies and trends in the mobile sector. Second, we assess the extent to which new measurement approaches are needed and how they can contribute to providing data-driven insights that meaningfully inform policymaking. Finally, we offer insights that provide guidance to policymakers on strategies for promoting a healthy measurement ecosystem to support good evidence-based policymaking.

In doing so, we seek to shed light on the following research questions:

1. What is the state of the measurement ecosystem (i.e., how did the supply and demand for measurements evolve)?
2. Is there a single best measurement approach/design?
3. How can we address the strategic use of measurements by different actors to achieve a workable measurement ecosystem that can address future challenges?

In addressing these questions, we offer three key insights.

[7] See, for example, Bauer et al. (2010, 2011, 2013, 2016, 2020), Bauer and Lehr (2018), Clark et al. (2014), Lehr et al. (2008, 2011), and Lehr (2022). See also Stocker and Whalley (2018, 2019) for an assessment of the different factors shaping end-to-end customer experience, the role of complex and diverse value chains, and challenges related to determining universal service targets for broadband.

[8] We focus on services provided by MNOs — the principal providers of the wide-area, mass-market mobile broadband services based on cellular networking technologies. These services differ from other types of mobile services that might be nomadic (e.g., end-users that move among different WiFi access points or are quasi-fixed-location when using the service) or portable (e.g., cable-free connectivity allows devices or usage to be moved locally as when an individual user in a home shifts among different end-user devices or when a portable cell site is moved to provide additional capacity in an emergency or for a special event). Historically, MNOs emerged as subsidiaries and complements to the network operations of fixed telephony operators, and the technologies used to deliver service have advanced through multiple generations of standardized technologies (from 1G to today's 5G) (e.g., Lehr et al., 2021). Today, the landscape of MNOs has expanded significantly in recent decades to include legacy cable television companies and a host of new MNOs and technologies that were not designed with MNO business models in mind. For example, many cable broadband providers rely on WiFi as their principal wireless technology, and newer satellite-based broadband providers and other developments in the broadband infrastructure sphere (connected vehicles, neutral hosts, etc.) are further expanding the range of business models and value chain relationships that may compete with and/or complement the operations of MNOs.

- *First*, the measurement ecosystem has become more capable and complex. Increased measurement capabilities are both enabled and necessitated by the expanding capabilities of our digital infrastructures. Regulatory policies will be needed to identify and intervene when tools are missing, or available tools are being misused.
- *Second*, there is no single best measurement source, nor should there be. Expanded measurement capabilities give rise to data-rich/information-rich environments and a variety of (new) asymmetric information problems. Different contexts will require different measurement approaches and designs. To address different information problems and policy questions, a larger number of industry players need to take and share different detailed data and measurements.
- *Third,* mobile broadband measurements are becoming more important and strategic. Incentivizing stakeholders to proper conduct will present a challenge as there are expanded opportunities and incentives to distort the measurement ecosystem. However, given the multistakeholder nature of the measurement challenge, and the fact that the capabilities to collect and interpret measurements will not be symmetrically distributed across industry participants, cultivating and fostering a third-party measurement ecosystem to supplement measurements by governments and network service providers is critical.

The balance of this paper is organized into three sections. Section 2 explains the changing nature of the measurement challenge and assesses mobile broadband measurement approaches, their capabilities, and evolution. Section 3 then shifts the perspective to three policy challenges and the lessons learned for a comprehensive and effective measurement ecosystem. Section 4 sums up and concludes.

2. Changing Measurement Ecosystem

Mobile broadband constitutes essential infrastructure to enable a wide range of innovative services and use cases, whose rapidly evolving nature gives rise to a set of pressing challenges for policymakers. In this context, it is important to realize that the requisite investment needed to provision (next-generation wireless) infrastructure will be mostly private and under the control of multiple (competing) industry stakeholders. Despite the fact that mobile performance measurements are increasingly available to inform strategic decision-making by these actors,[9] aggregating and merging data collected via a range of methods at different vantage points by different actors, and making sense of them to gain end-to-end insights, is complex.

The complexity in broadband measurement can result in challenging decision-making issues with ambiguous implications for measurement collection, processing, aggregation, and reporting. For example, as the increasing diversity of measurement sources utilizing different methods involves different actors and is motivated by different goals, the insights they

[9] As we further elaborate below, the collection, use, and sharing of broadband (performance) measurement data will be inherently strategic.

provide might not be aligned.[10] Indeed, if there are good but diverse measurement sources for the same question (e.g., "is MNO 1 or MNO 2 providing better service?"), this may result in measurement dispersion.

The measurement challenges can be explained in more detail by considering the demand and supply for measurements, both of which are changing. On the one hand, the need to measure more things on a more granular basis and to accommodate and reconcile measurements from a greater range of new and legacy stakeholders with strategic interests that are not aligned is growing. This is creating a more dynamic, complex, and multi-stakeholder decision environment. On the other hand, the technical capabilities and ability to measure more things and share those measurements and make use of them on a more fine-grained basis are larger than ever before.

2.1. The growing complexity of measuring mobile broadband performance

The nature of the measurement challenge and its complexity have changed and will continue to evolve. Not only is the measurement challenge increasingly complex, but it is also multi-stakeholder in its basic nature. Consider the Quality of Experience (QoE) on an end-to-end basis and as perceived by end-users. It is composite of QoS across multiple dimensions (latency, jitter, data rates, availability, etc.) and multiple components under the control of multiple stakeholders (including end-users and potentially with opposing interests). The resulting complexity reflects the contractual and non-contractual relationships between relevant stakeholders on the one hand, and the variety of space, time, and service contexts, on the other.

Mobile broadband is not just about telephone companies and telephone services, but a wide array of network services supporting a much larger and more demanding array of applications. Both the range and the diversity of QoS/QoE requirements have expanded, and the dimensions along which QoS dimensions may vary more dynamically has increased—they have become much finer-grained in space, time, and context. Additionally, the potential burstiness (peak-to-average-ratio of data rates) of specific applications has greatly increased. This can be seen most easily by considering that the range of applications and performance variation possible when the maximum data rate is 1Mbps is much smaller than when the maximum data rate is 1Gbps or 1Tbps.[11]

While this variety suggests that there is no single measurement perspective that is uniquely correct or worth considering, any aggregation or averaging to summarize performance in any dimension involves weightings that are potentially 'value-based'. For example, determining what constitutes 'fair' access or average performance depends on what the average is computed over. Generally speaking, the measurements we want may be QoE, but the measurements we will have will be multiple QoS/QoE from different stakeholders with

[10] Even the class of reasonable measurements may depend on what question is being asked. For example, the assessment may vary when comparing answers to the question "can I make a telephone call?" compared to the question "is a telephone call made from this street corner better if made via MNO 1 or MNO 2?".

[11] Note that while the range of nodes capable of supporting such data rates is quite limited today, it is increasing over time.

different vantage points that may all be valid, but are not necessarily aligned. In most cases and contexts, QoE will be an end-to-end composite of multiple QoS that may be associated with different metrics aggregated from multiple sources that may themselves reflect different goals and objectives.

It thus seems obvious that—in addition to advanced measurement approaches and the disclosure of verifiable measurement data or insights—we need an 'algebra' to understand how to composite different QoS factors to derive useful QoE insights. Making sense of the individual contribution of each component to overall end-to-end performance metric—as well as reliably identifying performance bottlenecks—is critical but challenging, especially for policymakers.[12] Moreover, as we laid out above, complexities emerge given the complex and evolving value chains that vary depending on a range of factors such as the service under consideration, the specific (local) context and network technology/topology, and time.

In addition to the general challenges that the evolution of the ecosystem poses for end-to-end and QoE network performance measurements, the evolution of mobile technologies adds increased complexity. Table 1 summarizes four inter-related factors that have transformed the economic and technical environment for mobile broadband networking and services.

Table 1: Identified Technical and Market Changes in Mobile Networks that are Restructuring the Nature of the Measurement Challenge

Factor	(Anticipated) Technical & Market Change
Small Cells	• Measurements more fine-grained and potentially variable on geo-localized basis (e.g., due to locally (co-)specialized network environments). • New business models and actors will factor into performance considerations (e.g., more customization (locally) and end-user involvement).
Spectrum Sharing	• More spectrum is shared across diverse regulatory regimes, between range of actors (e.g., among service providers or between service providers and end-users) on a more dynamic, geo-localized basis.[13] • More fungibility across bands due to apps/networks being more frequency agile.
Artificial Intelligence/ Softwarization	• More of (near) real-time and dynamic network management, pricing (transactions), and other modifications will be automated because of ▪ time (humans cannot react fast enough), ▪ complexity (bounded rationality constraints), ▪ security/reliability (eliminate human errors, shift liability for control, risk management/predictive and anticipatory measures).
Cross-layer Optimization	• Core network resources (spectrum, compute, data-layer connectivity, security risks, etc.) will need to be jointly optimized across layers (from Layer 1-7) and ultimately to analog ecosystem (regulations, business models, etc.) that OSI layers[14] are nested in.

[12] We will discuss this in more detail in Section 3.3.

[13] See, for example, Lehr and Stocker (2023).

[14] The OSI (Open Systems Interconnection) reference model is a conceptual framework that standardizes the functions of a telecommunication or computing system into seven distinct layers. It was developed by the International Organization for Standardization (ISO) to facilitate communication between different systems and promote interoperability. Each layer in the OSI model is responsible for specific tasks and provides services to the layer above and below it.

	• Even if one restricts attention to Internet layers 1-3, cross-layer optimization is required

Source: Authors

To tackle the challenges posed by limited radio spectrum, spectrum needs to be reused more intensively. This has given rise to a shift to *smaller cell architectures* for wireless networks, wherein each cell site provides service over a smaller geographic area, which additionally facilitates the use of more spectrum resources, in particular, previously untapped higher frequencies.[15] Whereas the capacity of individual cells can thus be expanded, many more cell sites and lots of additional investment to integrate those smaller cells into service provider networks are required.

Additionally, *spectrum sharing* among users and networks is also increasing to further facilitate spectrum reuse. Managing all this complexity requires more intelligent software applications and orchestration, facilitated by *softwarization* of information technologies[16]. To fully realize the capabilities associated with the 5G design requirements (ITU-R, 2015) will require more complex network management software systems (e.g., Network Function Virtualization, NFV), and indeed, *Artificial Intelligence* (AI). The goal for 6G standardization efforts is to enable AI-native network environments to support such resource-demanding applications as Augmented Reality (AR) and Virtual Reality (VR) and enable dynamic resource provisioning capabilities. These AI applications are data-hungry and empower context-aware automated control of key network functionality (e.g., fault recovery, dynamic provisioning, and interaction with other network components) as well as improved interoperability with end-users and human operators (e.g., by presenting human-readable summaries of machine-readable measurements via data visualizations and other AI-processing techniques).

Finally, in these increasingly complex technical environments, ensuring QoS meets intended levels will make use of *cross-layer optimizations* involving the tuning and selection of protocols across multiple networking layers. For example, the right transmission protocols to use depends on the frequency used so the selection of Radio Frequency (RF) resources will require changes at higher protocol layers and along the end-to-end path. For example, the delivered quality of a streaming video can be sustained by using more spectrum (potentially available at a higher frequency) or by using more computationally intensive coding schemes, and the choice of which coding scheme, spectrum resources, or other network resources to use may vary with the type of content or application and its tolerance for delay, bit errors, and other lower-level technical performance metrics. The ability to automate and take advantage of different ways to provision and support dynamic services gives rise to a need for more complex and fine-grained measurement capabilities.

[15] Higher frequency spectrum has different propagation characteristics including need for line-of-sight (LOS) connectivity between the base station and the receiving radio device. For smaller cell architectures, this LOS limitation is less of an issue.

[16] Softwarization refers to the movement of functionality from dedicated hardware into software, which enables it to be flexibly modified on a faster timescale and facilitates unbundling the location of where actions take place and their control, or the delocalization of functionality and virtualization. Virtualization allows higher-level applications to share lower-level infrastructure and have customer-specific experiences (e.g., Shukla and Stocker, 2019).

2.2. Performance Measurements for Mobile Broadband Services

Historically, network availability and performance measurements in mobile communications have relied on three approaches. Whereas *unattended probes* and *drive/walk tests* constitute more traditional approaches, in recent years, the supply of mobile network performance measurements has significantly transformed. The way and the type of data collected have notably changed with novel data collection capabilities enabled by smartphones. The historical approaches to mobile performance measurements are being replaced, or at least complemented, by new *crowdsourced measurement approaches* using mobile applications installed on smartphones. As of today, the legacy approaches of probes, and drive/walk tests coexist with the new crowdsourced methodologies. Table 2 provides an overview of the different measurement approaches.

Table 2: Measurement Approaches

Approach	Description
Unattended Probes	• Deployed in fixed locations to monitor network performance. • Probe-based monitoring systems use equipment deployed in fixed locations that are intended to be representative of the environment of residential customers. • Geographical location requirements are set to comply with specific technical conditions. • The volume of measurements (number of probes and frequency) is determined by the data collection strategy.
Drive Tests/ Walk Tests	• Measurement equipment (specific hardware and software) is mounted in cars (or carried by pedestrians). • Measurements are collected according to a pre-defined measurement methodology and plan across a geographic area.
Crowdsourced Measurement Approaches	• Measurements are taken by individual end-users and contributed to large databases.

Source: Authors

Arguably, the systematic approach of probes and drive tests[17] attains high measurement reliability since respective measurement plans are designed to control for the impact of factors that cause variability in the measurements, such as signal fading or network load conditions. Drive tests adhere to standardized methodologies, ensuring consistency and enhancing the comparability of results. These standardized approaches provide guidelines and procedures for conducting measurements, allowing different operators or researchers to obtain results that can be easily compared and analyzed, making drive testing well suited for benchmarking. Moreover, drive tests are conducted in a known context. This includes the information of the measurement conditions as well as the equipment used, which is usually professional tailored equipment. The use of this professional tailored equipment enables access to additional information of the state of the network that allows in-depth analysis of network performance, including signaling messages, handover processes, etc. Having access

[17] For the sake of simplicity, we will in the following refer to the category of "drive tests/walk tests" as "drive tests".

to this information helps identify specific issues or anomalies in the network and facilitates troubleshooting and optimization efforts.

Despite these strengths, drive testing is very resource and labor-intensive. It requires dedicated test vehicles, specialized measurement equipment, and trained personnel to conduct the tests. The process involves extensive planning, coordination, and data collection, which is also time-consuming and costly. For this reason, drive tests have limitations in terms of scale and scope. Conducting drive tests across a large geographic area or nationwide coverage can be expensive and logistically challenging. It is not feasible to cover every location, and therefore, the measurements may not provide a comprehensive view of the entire network.

Moreover, although drive tests can deliver precise information and valuable insights about the radio access network, they are more limited in assessing end-user QoE for several reasons. First, the tests are typically performed using specific test devices and predefined test scenarios, which may not accurately represent the diverse range of devices and usage patterns of actual users. Additionally, the complexities of the Internet architecture may require a much larger number of tests than physically feasible with drive tests to assess QoE for specific applications. Finally, drive tests capture network performance during a specific time window, which is no problem for some measurements, such as signal strength, but may not capture the full variability of network conditions, potentially leading to incomplete or biased measurements in response to network load variation.

In contrast to drive tests, unattended probes can collect lots of measurements and capture the influence of network conditions variations as they provide continuous monitoring of the mobile network performance at fixed locations. This continuous monitoring allows for timely identification of network issues and performance degradation. The most obvious strength of this approach is that the probes operate autonomously and collect data without the need for manual intervention, which makes them easier to scale as they are not so reliant on human labor.

On the other hand, unattended probes are typically deployed at fixed locations within the network. This limited location flexibility may result in uneven location coverage and may not capture performance variations in all areas, especially in remote or dynamically changing environments. Moreover, while unattended probes provide valuable network-level insights, they may not capture the full user perspective or experience. They mainly focus on technical network parameters and do not reflect the actual customer experience. Aspects such as user behavior, device-specific issues, or location-specific conditions may not be adequately captured by unattended probes alone.

In contrast to drive testing and probes, crowdsourcing allows for scalable measurement collection with both high spatial and temporal resolution over broad areas, thereby enabling the collection of a volume of measurements that would be prohibitively expensive to do with drive tests or fixed-location probes. This extensive coverage provides a more comprehensive understanding of network performance, including both urban and rural areas. Moreover, crowdsourced measurements can better reflect the actual user experience of mobile network services because they draw on data from a diverse range of devices and users, with different approaches, as described below in Section 2.4.

However, while crowdsourced methodologies for mobile performance measurements offer several advantages, they also have some inherent weaknesses. Crowdsourced measurements suffer from many sources of potential biases, as they heavily rely on user participation and their willingness to provide data. This can lead to an uneven representation of user demographics, geographic areas, or device types, potentially skewing the results and limiting the generalizability of findings. Additionally, unlike in laboratory or drive test scenarios, in crowdsourced measurements there is limited control over the measurement environment, which usually requires additional post-processing for context inference that can subsequently help isolate specific variables and assess their impact on network performance. Similarly, as the mobile operating systems (OSs) do not expose all network information to third parties, the network conditions of the measurements cannot always be precisely identified. Finally, crowdsourced data often require sophisticated analysis and interpretation techniques. The sheer volume of data collected can pose challenges in extracting meaningful insights for which sophisticated post-processing (possibly involving the use of AI) are necessary to derive actionable information and meaningful insights.

Depending on the questions sought to explore with the measurements, either approach may be more appropriate. Thus, if very precise performance measurements are needed for a specific situation (e.g., capacity planning for a small cell) then a fixed unattended probe or drive test measurement design may be best; whereas if data is needed to characterize the average performance experienced over the entire coverage footprint of a mobile network over time, then crowdsourced measurements may be best. And, in many contexts, being able to combine the measurements from multiple sources may offer significant advantages[18]. Table 3 provides an overview of the strengths and weaknesses for each of the measurement methods described above, thus providing the basis for the following sections in which we describe how the mobile measurement market and the respective measurement approaches have evolved. We will explain the rising importance of crowdsourced methods to better reflect

[18] For example, in evaluating the performance of broadband services in the U.S., policymakers rely both on crowdsourced measurement data from Ookla, as well as fixed-probe measurement data from the SamKnows Measuring Broadband America program, indicating the complementary nature of the measurement approaches (potentially across multiple dimensions related to the technical method, spatial/mobility matrix and resolution, etc.). Ookla uses drive testing tools (Ookla Wind) on top of their crowdsourced measurement tools (see Ookla, n.d.-b). For a discussion of Measuring Broadband America (MBA) program that was launched with SamKnows probes see Bauer et al. (2010) or Sundaresan et al. (2011). As of August 1st, 2023, the FCC announced that it will no longer be using the SamKnows probes to collect MBA data (see FCC, 2023). The FCC will continue its MBA efforts, but it is unclear how it will measure fixed broadband performance. The decision to no longer use SamKnows may be related to the recent announcement that Cisco is planning to acquire SamKnows (Salter, 2023). The MBA SamKnows data had the advantage of using fixed probes that isolate performance measurements to focus on the service provided by the participating ISPs using standardized measurement tools; however the number and location of SamKnows probes leaves many broadband access connections and service providers service unmeasured. The Ookla crowdsourced broadband measurement data is described at https://www.speedtest.net/ (Ookla, n.d.-c) and also in Connelly (2021). The Ookla measurements rely on end-users voluntarily connecting to Ookla's cloud-based measurement infrastructure and provides greater flexibility in the timing and scope of measurements than are enabled by SamKnows-style measurements, but also provides less control over the measurement context and hence control over such factors as the quality/performance of a customer's home network.

spatial performance than probe-based methods, which is essential for evaluating mobile services.

Table 3: Comparing Measurement Approaches: Strengths & Weaknesses

Approach	Strengths	Weaknesses
Unattended Probes	• Unattended (automated) • High temporal resolution • Other strengths similar to drive tests/walk, except for mobility	• Fixed locations (limited spatial representation) • Limited user perspective
Drive Tests/ Walk Tests	• Standard methodologies are used, which enhances consistency and comparability of 'local' results • Known context (including measurement equipment) yields high measurement reliability • Availability of the entire protocol stack information (L3)	• Resource and labor intensive • Limited scale (expensive) and scope (no country-wide coverage) • Low representativeness of end-user QoE • Limited temporal window
Crowdsourced Measurement Approaches	• High spatio-temporal resolution • Broad scope (w.r.t. spatial footprint and range of applications) • Cost efficient and scalable • End-user QoE KPIs can be measured	• No standard methodologies • No control over the end-user equipment (e.g., device and home network; hardware, use, and configuration) and context, potential biases • No availability of full protocol stack information • Complex data interpretation

Source: Authors

2.3. Evolution of Drive/Walk Tests

Since the early days of mobile networks, drive tests have been the traditional approach for measuring performance in mobile networks. As the predominant service in the early mobile generations was voice, drive tests focused on (i) measuring radio coverage, and (ii) assessing the QoS of voice calls. The radio coverage was measured using standardized GSM measurements collected by the user equipment (UE).[19] To measure voice call quality, the most common Key Performance Indicators (KPIs) were the call drop rate, call set-up failures, and voice quality. The voice quality was assessed both technically and subjectively.[20]

---

[19] The principal metrics were RxLev (received power from beacon carrier or dedicated channel) and RxQual (estimated bit error rate before channel decoding in the dedicated channel) (Mouly & Pautet, 1992).

[20] For example, Karkhanechi and Soderstrand (1997) describe how metrics like signal strength, signal to noise ratio, bit error rates, and other technical metrics may be used to assess voice quality in cellular mobile phones. For subjective measurements of voice quality, the ITU standard P.862 defines a "MOS" (Mean Opinion Score) standardized approach for assessing the perceptual evaluation of speech quality (PESQ) (see ITU-T, 2001). With the rise of VoIP calling, a number of third-party providers offer software to assess the quality of voice calls over different platforms (see, e.g., Lamberti, 2022). Moreover, with the rise of high-quality voice services and proliferation of new software models for automated and human-speech interactions have expanded, more advanced psychoacoustic measurement capabilities have become available from measurement providers like Rhode & Schwarz. For example, ITU-T P.863 offers a standardized approach for evaluating how the received voice signal compares with a reference signal for testing the quality of mobile-to-mobile calls that is intended to capture human hearing psychoacoustics (see, e.g., Rohde & Schwarz, 2023). As this brief discussion illustrates, the need and optimal design for network measurements, optimal metrics and their interpretation are inherently context-specific—and this is true for all components involved in determining end-to-end QoE.

With the inclusion of data services in the 3rd mobile generation (3G), coverage-related measurements were still used, specifically RSCP (Received Signal Code Power; on the pilot channel transmitted by the base station) and E_C/I_0 (signal-to-interference-plus noise ratio on the pilot channel), but in addition uplink (UL) and downlink (DL) data rates began to be tested as they were deemed as the primary proxy for the quality of the data service, which was then fundamentally used for web browsing.

As UL and DL speeds increased with 4G services and smartphone adoption, the diversity of mobile apps and services expanded rapidly in terms of the range of data rates and QoS requirements. At one end, services like voice and SMS were relatively undemanding in terms of network performance/QoS as reflected by data rates and latency. Applications like video streaming and video conferencing, however, were much more resource intensive and the dimensions along which QoS issues might arise expanded.[21] In addition to coverage measurements such as RSRP (Reference Signal Received Power) and RSRQ (Reference Signal Received Quality, similar to a SINR), most drive testing equipment introduced video speed testing. The tests typically allow the 'user' to configure a set of automated calls to assess the QoS and the QoE/customer experience. The testing not only enables measuring basic connection parameters such as bit rate or latency but is also augmented by more user-centric and audio- and video-specific quality scores computed with standard algorithms.[22]

It is important to note that international standardization bodies such as the ITU-T (International Telecommunication Union — Telecommunication Standardization Sector) and the ETSI (European Technical Standards Institute) have established reference standards for implementing drive testing. The use of those standardized methods facilitates the sharing and (subsequent) evaluation of results. While the use of such standardized methods is widespread, it is not universal. However, there are examples of regulations that mandate the use of specific standardized measurement methods.[23]

The capabilities and results of drive tests can vary across different implementations and designs. They are influenced by decision-making at three different levels involving different

[21] For example, with SMS, the challenge is to deliver a relatively short text message within a relatively short period of time. However, delays in delivery measured in minutes, while unpleasant, do not doom the service (and may in some cases not even significantly impair customer experience). With video conferencing, things are different. The timeliness of packet delivery, and network performance and QoS (e.g., latency and jitter) more generally, are much more critical. There is a broader range of QoS impairments that may be tolerated while still rendering the service usable. For example, dropped video frames, reduced resolution, or even failure of the video may still allow participants to continue a video conference call productively. However, the subjective evaluation of performance and the options available to end-users, MNOs, and third-party application providers (e.g., Zoom) to impact the customer experience are richer than for SMS or other simpler applications. For example, for some video calls, the video is essential whereas for others it may be a nice-to-have additional feature that is not very important to the users. The diversity of electronic communication options (video call, voice call, text messaging, chat, email, etc.) allow end-users to select among a wider array of service options.

[22] Beyond such standard algorithms, the linking network performance/QoS and QoE/customer experience has been an active research field. For an excellent overview see, for example, the edited volume by Möller and Raake (2014).

[23] See, for example, Section 3 for reference to QoS regulation in the context of the EU's 2018 European Electronic Communications Code (EECC).

stakeholders: standardization bodies, measurement equipment manufacturers, and measurement campaign designers.[24] Standardization bodies specify technical aspects and recommend good practices that allow the measurements to be comparable. Measurement equipment manufacturers (or measurement tool vendors) usually incorporate these recommendations and amend them with further details and capabilities. Lastly, the designer and promoter of the measurement campaign (e.g., the MNO or NRA) defines the specific approach to be used, which is then carried out, usually by a contractor company. Before we provide more detailed explanation, Table 4 provides a high-level overview of the methodology specification.

Table 4: Methodology Specification: The Case of Drive Tests

Level/ Stakeholder	Explanation
Standardization Bodies	• Promote reference standards to evaluate audio and video quality • Publishes best practices/recommendations to undertake drive tests
Measurement Equipment Manufacturers	• Special purpose hardware and software for data collection and analysis • QoS assessment capabilities based on standards • Non-standardized QoS assessment capabilities for popular Internet services
Measurement Campaign Designers	• Design according to own criteria and purpose • Measurement capabilities choices to match available budget

Source: Authors

Standardization bodies. Standardization bodies, such as the ITU or the ETSI have long contributed to characterizing network performance and end user perceived quality from different angles. One of their most important contributions has been the development of standards to evaluate video and audio quality, whose methods and approaches have undergone notable transformation over the last few decades to adapt to technology changes. During the legacy circuit-switched networks' days, these international organizations promoted reference standards to measure voice and video quality and assess subjective components associated with quality perception (ITU-T, 2001; ETSI, 2007). Additionally, the ITU provides best practices to harmonize criteria for the design and execution of drive tests campaigns (ITU-T, 2019)[25].

Measurement equipment manufacturers—Tool vendors. Drive tests are carried out using specific measurement tools designed for network testing and optimization. Some of the most popular drive testing equipment is developed by the companies Rohde & Schwarz, Keysight technologies, Anritsu Corporation, and Infovista, among others. This equipment comprises both specific hardware and software to perform the measurements, which is designed to provide access to all the information[26] available to the deployed mobile devices for the

[24] Measurement campaigns may be designed by the interested party commissioning the campaign (e.g., MNOs or NRAs) or by the contractor company doing the measurements.

[25] ITU-T E.806 describes best practices for QoS measurement in mobile networks, including the monitoring systems characteristics, post processing recommendations, and sampling methodologies on a technology-neutral basis.

[26] Including layer-3 messages of the radio interface, although sometimes access to some information may be hidden by chipset manufacturers and therefore might not be available either with some drive tests equipment.

operation of the communication with the network base stations. This allows accessing useful information of the network state for high-accuracy tasks such as protocol debugging and parameter optimization.

The drive test equipment options and capabilities have undergone notable changes to adapt to the changing nature of the measurement challenge. For example, early drive test equipment relied on a mobile terminal connected to a laptop, which along with the absence of GPS, restricted the pace of measurement recording in both time and space. With the advent of the smartphone, some of the measurement products in the market (particularly the lower end) use special-purpose firmware running on general-purpose terminals, but most of the equipment continues to be professional and usually hardware-specific. Radio scanners are broadly used to obtain high-quality RF measurements that may be used to infer higher-value information and insights[27].

Finally, vendors now usually incorporate quality assessment capabilities, usually based on the ITU or ETSI standards described previously. Some go one step further and include more specific, non-standardized measurements for popular applications related to cloud storage, or social networks, among others. For example, some measurement tools (such as Rohde & Schwarz's Qualipoc) can automate the test of a typical Facebook session (SwissQual, 2020b). Not surprisingly, it has been found that the load balancing algorithm applied by Facebook to distribute the requests among their servers can have a significant impact on performance indicators such as picture upload times.

Measurement campaign designers. As mentioned above, drive tests are usually commissioned by MNOs or NRAs, but designed and executed by smaller contractor companies usually providing services in other infrastructure markets (utilities, transportation, etc.), as they require on-site human resources and the use of professional equipment from the large global technology companies described above. As per their own criteria and purpose, they may also make choices that influence the results.

2.4. Evolution of Mobile Crowdsourced Measurements

The programmability of mobile OSs, particularly Android, has played a crucial role in enabling the massive collection of Mobile Crowdsourced (MCS) measurements. However, different stakeholders have adopted varying approaches to data collection. There are two primary approaches of collecting MCS measurements[28], each one having distinct

[27] For example, scanners allow to simultaneously test several base stations, operators, frequencies, and technologies, and thus they are broadly used for benchmarking. In addition to providing physical-level measurements, the scanner can also include decoding of downlink channels. This allows some vendors to use the scanner for more advanced measurements, such as estimating an LTE base station load based on decoding the scheduling messages broadcasted on the downlink. See, for further reference, SwissQual (2020a).

[28] In addition to the two approaches presented, there are other network measurement applications that provide detailed data about RF coverage information broadcasted by the mobile network. Examples are Reference Signal Received Power (RSRP) or Reference Signal Received Quality (RSRQ) for 4G and 5G networks. The technical orientation of these applications, however, renders them usable by a relatively small group of end

implications for the methodology and interpretation of the results and revealing different trade-offs.

Speedtests (user-initiated). Speed tests are the oldest technique for network performance data collection. Their use was initially motivated by end-users' desire to verify that the broadband 'speed' they were getting corresponded with what was advertised. Speed tests were first used in fixed networks through a web client, but they became prevalent in the form of mobile apps as mobile broadband penetration increased. In general, the speed testing service is provided for free to the user, in exchange for collecting the results along with other relevant data on the network performance, which the app provider may use for its business, e.g., consulting services. The most popular speed test service is probably Ookla's, the Seattle-based company delivering the first speed tests that owns the domain *www.speedtest.net* and the registered trademark 'Speedtest'.

The speedtest approach implies active measurements, the measurement is user-initiated, and the goal is to overflow the network with data request to test its maximum capacity. These applications run in the foreground on end-user mobile OSs. As the goal is to measure access network capacity, the tests are performed against local servers located as close to the users as possible to avoid the complexity of the Internet topology affecting the results.[29] Because in practice it is not possible to locate servers close to all potential end-users due to cost reasons, this approach has evolved in recent years by performing multi-sever testing to identify unreliable results of the access network performance if the measurements towards different servers provide substantially different results.

The fact that the test is user-initiated poses several challenges for data collection. First, this approach introduces potential biases, as users are more likely to initiate a speed test when either they have changed their residential or work location (e.g., they have moved to a new house, city, or company) or subscription, or they are experiencing connectivity problems. This affects the spatio-temporal resolution of the collected data. Second, as the test is initiated by the user, the sample size that may be collected is limited. This reduced sample size hinders proper statistical analysis of the measurements and the results' significance. To address this problem, approaches based on Software Development Kits (SDKs) emerged.

Software Development Kits. SDKs are installable software packages for third parties to use in a particular framework, namely, a mobile application not related to mobile performance measurements. Such apps could relate to transportation, restaurant bookings, social networks, or any other. To overcome the problem of user-initiated measurements, some companies have

---

users with a technical background. This, in turn, limits the volume and scope of data generated. Broadly speaking, the applications allow to store and export all data collected by the end user so that they can further process them. Some of the measurements from these apps are contributed to open data projects on mobile infrastructure, such as Mozilla Location Services (Mozilla, n.d.) or OpencellID (Unwired Labs, n.d.).

[29] If the server were located far away, other elements involved in the service delivery (e.g., an interconnection point or an adjacent or remote network) might become the bottleneck determining the maximum connection capacity and hiding the performance of the targeted (radio) access network (see also, for example, Stocker & Whalley, 2018, or Feamster & Livingood, 2020). Thus, an extensive (own) server network is an essential asset for this approach and critical for determining the meaningfulness of the insights gained through the measurements.

developed SDKs to embed the measurement-collecting code into other applications with broader use. In this way, much more data is collected, and the circumstances of the measurements can be better controlled to minimize biases, as the measurements are taken in the background on the end-user mobile OS.

SDK-based measurement approaches have enabled sophisticated MCS business models, as third-party app developers need to be incentivized to use the SDKs and collect data that are beyond their own business-related data needs. Thus, in most cases, app developers receive payments from SDK producers in exchange for the measurements they provide. This approach was pioneered by Tutela, a company acquired by Comlinkdata in 2019, but has been followed by others afterwards. One of the largest companies collecting MCS measurements with this approach nowadays is OpenSignal[30], which was also acquired by Comlinkdata and whose initial approach was focusing on speed tests.

These two measurement approaches are not mutually exclusive, and most MCS data companies currently combine them in their data collection given the complementary information that they provide. However, they usually lean towards one of them to develop their value proposition and business strategy. Table 1 in the Appendix compares the methodologies employed by the two leading MCS performance measurement companies, Ookla and OpenSignal. Broadly speaking, whereas Ookla relies on speed tests run by end users, OpenSignal draws on measurements taken via their SDK. However both combine active (user-initiated, running in foreground) and passive measurements (data collection running in background).

Table 1 in the Appendix provides a detailed description of the approaches of the two larger MCS providers for comparison and illustrates the several sources of differences in the approach followed by the companies. The primary sources of differences can be grouped in the following key aspects:

- **KPI definitions and implementations**. Companies create their own KPIs with their own definition and range (e.g., 0-10, 1-1000), such as Coverage Score, Speed Score, Reach, Coverage experience (see Table 1 in the Appendix for further reference). They are mostly determined by the crowdsourcing method and its limitations rather than by their specific approach (e.g., it is difficult to determine coverage in places with no reported measures). Companies disclose general definitions of their KPIs and describe which factors may influence the results, but not how the KPIs are specifically computed (its implementation). As there are large volumes of spatio-temporal data, there are usually many potential ways to compute the raw values to obtain simple metrics.

[30] The company OpenSignal signal originated in the United Kingdom with the aim to empower end users to verify which MNO provided the best service in those areas they visited more frequently. Its initial approach was speed test-based, but it has progressively shifted to SDK-based data collection, although it keeps its speed test mobile app. This change has led to consolidation of the sector as exemplified by the following mergers: Comlinkdata acquired Tutela in 2019, who pioneered the SDK approach, and more recently OpenSignal in 2021.

- **Summary statistics choice.** The need to reduce large volumes of data to representative KPIs requires companies to make choices regarding their statistics, for example, including the use of median, mean, percentile values, or even combinations thereof. For example, Ookla computes 1:8:1 weight of 90, 50 and 10 percentiles for Speed Score, and uses median values for access speeds, whereas OpenSignal provides average values (Connelly, 2021; OpenSignal, 2023).
- **MCS approach.** Different measurement approaches have a strong influence on the results, as they may be measuring different parts of the network. For example, whereas Ookla only relies on passive measurements taken in the background of their app for coverage data (e.g., received signal power, RSRP in 4G), OpenSignal collects the majority of their data via passive measurements from their partners, including speed, quality, and application experience measurements (OpenSignal, 2023). Therefore, the speed measured by OpenSignal is not the maximum achievable speed of the channel (as it is with user-initiated speedtests, which aims to overflow the network with data request), but the actual speed that end-users receive when using their smartphones' services and applications. Moreover, contrary to Ookla's approach to local server testing, OpenSignal tests against the CDN (Content Delivery Network) servers of leading companies like Google or Amazon. The rationale is that CDNs deliver a significant fraction of the user's total traffic and therefore the data rates towards these is highly indicative of the actual data rates (i.e., 'speed') that shape end-users' QoE.

It needs to be noted that the lack of control over how, when, where and by whom the measurements are taken means that sophisticated data post-processing techniques are needed to minimize potential biases. Moreover, the measurements are collected through the APIs of the prevailing mobile OSs, namely, Android and iOS. The network information accessible through these APIs is much more limited than that collected through network measurement-specific devices used in drive tests. For example, common variables for signal strength and signal quality (e.g, RSRP and RSRQ in LTE) are available through smartphone APIs, but not more detailed network information like the System Information Block (SIB) (Dahlman et al., 2013) that provides parameter configuration of the mobile network and therefore allows to obtain more detailed information and more valuable insights. Finally, the data collection techniques and the intensive data post-processing that MCS approaches need to minimize their potential biases are proprietary and one of their key intellectual and industrial property assets, which make MCS performance measurements hard to compare and may add some opaqueness.

However, unlike drive tests, MCS measurements scale well, as their collection and processing are carried out primarily by automated means. Thus, country-wide statistics are available, which would be economically unfeasible with drive testing methods. The large volume of data means that, despite using more heterodox statistical methodologies, the representativeness of the data may be very good, especially in urban areas, where population density (and therefore sample volume) is high. MCS approaches allow for higher spatio-temporal resolution than drive testing, which may be relevant for some policy and industry problems, as described below in Section 3. Finally, MSC methodologies allow testing that informs end-to-end QoE as perceived end-users, in a way that drive tests cannot. Background tests running on end-user smartphones allow monitoring the actual user experience when using a wide range of applications and services, such as video streaming, video calling, gaming, etc. In contrast, drive tests may only test specific speed and quality metrics (i.e., data rates and latency/jitter) against local servers.

2.5. The State and Future Trajectory of the Measurement Ecosystem — Reflections and Discussion

In the past, radio coverage was the focus of mobile performance measurements because it was the primary factor driving the QoS for voice services provided by vertically integrated mobile companies. Radio coverage determined the reach and reliability of voice calls, and measuring it was essential for assessing the performance of mobile networks. However, the landscape of mobile communications has significantly evolved over time. The advancements in technology and the introduction of data-centric services, such as the mobile Internet and the broad range of content and applications that can be accessed, have transformed both the priorities and challenges in the mobile measurement domain. While radio coverage remains important, it is no longer the sole determining factor for assessing the performance and QoS of today's mobile networks. Several key factors have contributed to this shift.

First, the advent of data-centric services, such as web browsing, video streaming, and other mobile applications, has shifted the focus from voice-centric QoS metrics (such as call dropped rate or voice connection quality) and radio coverage metrics (such as signal strength and signal quality) to a broader range of network performance/QoS and customer experience/QoE metrics. These metrics include data rates, latency, packet loss, and application-level QoE metrics, as mobile connectivity is merely a necessary but not a sufficient condition to indicate customer experience (e.g., Stocker & Whalley, 2018). As mobile networks have evolved to accommodate higher data rates and lower latencies and a wider range of applications, the network architecture has become increasingly more important, as multiple actors and perhaps networks interact to jointly provide the service to the end user. All measurement approaches have adapted to this shift, but in different ways. Some have incorporated testing methodologies that target popular CDNs whereas others have employed multi-server tests.

Secondly, recent investments in expanding the capacity of the radio access network (RAN) have effectively addressed the surging demand for mobile traffic, thereby alleviating the pressure on the edge mobile network, which was previously often the most restrictive link in the service delivery chain. The deployment of high-density cells—including fiber-fed small cells and heterogeneous networks, along with the integration of expanded spectrum portfolios and advanced spectral efficiency of mobile technologies—has significantly enhanced the network’s capability to handle greater data traffic. As a result, the performance of the RAN has substantially improved over the last two decades.

Another key change of the mobile measurement ecosystem during the last decade has been the notably increased volume of measurements and their increased spatio-temporal resolution, primary driven by the emergence and development of crowdsourced methodologies. In the spatial domain, performing mobile tests from end-user devices have notably increased the spatial resolution, although with unavoidable disparities in measurement capabilities across geographic regions, with denser areas benefiting more from this approach. In the temporal domain, the adoption of background data collection through SDKs has facilitated the collection of a much larger volume of data.

The increased measurement capabilities of crowdsourced methods come at the cost of losing full control of the measurement conditions. Although these methods sacrifice control and may introduce some bias, the valuable insights gained from increased resolution often outweigh the limitations. All measurements methods have inherent noise. While traditional drive tests offer more control over measurement conditions, crowdsourced methods can deliver better results with sufficient data and post-processing, leveraging on data from smartphones' sensors to infer the measurement context. The fit and effectiveness of each approach depends on the specific context and objectives.

Drive tests allow for controlled measurements but may not capture the full diversity of network conditions and temporal dynamics, whereas crowdsourced methods provide a larger volume of data but with potential biases that need to be minimized. Choosing the appropriate method depends on the desired outcomes and trade-offs between control and scale.[31] Thus, if very precise performance measurements are needed for a specific situation (e.g., capacity planning for a small cell), then a fixed probe or drive test measurement design may be best; whereas if data is needed to characterize the average performance experienced over the entire coverage footprint of a mobile network over time, then crowdsourced measurements may be best. And, in many contexts, being able to combine the measurements from multiple sources may offer significant advantages.

However, the interpretation of mobile measurements has become more complex for several reasons. First, due to the inherent noise (uninterpretable measurement variation) present in all measurement approaches. Second, the tremendously increased complexity of the RAN, with multiple technologies, frequency bands, and cell sizes have become more intricate measuring mobile performance. The RAN has become a complex system with its own behavior, whose understanding increasingly requires leveraging advancements in data analytics and machine learning techniques. State-of-the-art research is exploring ways to understand the interactions and dynamics within the RAN to uncover hidden patterns, optimize network performance, and improve the overall user experience. Hopefully these breakthroughs will allow us to obtain meaningful and actionable insights from the increased measurements capabilities provided by crowdsourcing approaches, but this is live research.

Despite the increased measurement capabilities, the current ecosystem faces challenges and limitations. On the one hand, the standardization of mobile networks focuses solely on the information exchange required for the proper communication between terminals and network

[31] Measurements are a function of all design characteristics and different designs have different properties with regard to aspects like bias/noise, coverage, or flexibility. Bias/noise issues exist with all approaches—drive/walk tests have a sample selection risk since designers can manipulate the location of test nodes (what tests are taken?), whereas crowdsourced approaches can introduce bias of mobile users (e.g., sample selection because all crowdsources congregate in few places). It appears that the control of noise (uninterpretable measurement variation) is easier with drive test and fixed probe test designs. Whereas the current methodology diversity and mix of measurement approaches has complicated data interpretation, the meaningfulness of single methods or combinations vary tremendously across contexts.

equipment, and the measurements conducted by the terminals are designed accordingly. The 3GPP standardization processes often neglect the importance of defining practical metrics for the effective operation of mobile service markets in the future and thus the technology definition limits the subsequent measurement capabilities. On the other hand, the measurement capabilities of crowdsourced methods depend on the information exposed by the OSs of mobile devices. The level of information exposure affects the granularity and accuracy of crowdsourced measurements. Therefore, the capabilities of crowdsourced methods are largely determined by the OS policies and features that govern data collection and access to relevant network parameters. This information exposure has become key to sustain a new more dynamic measurement ecosystem, and this may become one key aspect of policy decision in the future.

Arguably, one may predict that the future of measurements will be characterized by the improved ability to do drive-testing at lower cost while also offering the benefits of crowdsourcing. In principle, this would yield better control and enable the launching of incremental active/passive measurements. We may anticipate that the proliferation of IoT/smartphones means potential vantage points are (nearly) everywhere and each could be loaded with software/applications that could take measurements on-demand at very low cost. While such a future implies improved measurement capabilities in the ecosystem, it does not necessarily imply better decisions. The relaxation of constraints on measurement (information) options that rendered drive-testing preferable to crowdsourcing in certain contexts brings with it new challenges that need to be considered (e.g., waterbed effects).

It is important to note that in principle, it is possible to replicate insights gained via drive tests with a large enough set of crowdsourced measurements or vice versa. In contexts where different approaches can yield the same insights, the choice of how one gets to any data set is an economic one, at least prospectively. When considering the cost-benefit trade-offs of alternative test strategies, it becomes important to consider that the economics of implementing different test strategies have changed over time. Moreover, due to irreversibilities (i.e., one cannot go back in time), the ability to forensically replicate a data set by different methods is limited. Consequently, determining what the best measurement approach is may differ based on whether one is interested in prospective or forensic (past) measurements. These differences may also factor into ex ante expectations about enforcement. For example, the decision by different stakeholders to collect or not collect data today that may allow ex post forensic prosecution for violations (that may be subject to Type I or Type II errors) can influence their ex ante incentives to support different measurement capabilities today. As we mentioned previously, this emphasizes that the choice and use of measurements is crucially shaped by strategic considerations.

3. Discussion & Implications for Policymaking

A healthy ecosystem for measuring mobile broadband availability and performance produces market intelligence and insights based on the combination of data from different actors, results from active and passive measurements, and includes a variety of differing methods reflecting different quality, provenance characteristics, and vantage points. In a world where value chains change dynamically across time, space, and context, different questions may require different measurement designs characterized by different trade-offs. It becomes obvious that a capable measurement ecosystem is diverse, combining a variety of measurements conducted by different ecosystem actors based on different methodologies in a flexible way. How to incentivize the collection of the data we want and need through 'good measurements' and ensure that data (insights) are shared, aggregated, and interpreted meaningfully is challenging, particularly due to potentially strategic behavior at all levels by all actors.

For example, strategic actors may disagree as to what the right choice is for measurement design among valid choices; or worse, they may advocate for measurement designs that are intentionally or accidentally invalid (misleading) choices. Whereas the absence of appropriate standards emphasizes the trade-offs related to the flexibility needed to facilitate context-specific measurement designs, bad measurements may be the result of coding or other implementation errors that may be difficult to detect. The latter may also be the result of intentional efforts to distort the evidence to induce decision-making outcomes more favorable to the sponsors of the bad measurement data. Bad measurements can result from intentionally selecting where to locate probes or what crowdsourced data to report (e.g., censoring data to exclude bad performance measurements which might be disguised as faulty corrections for measurement errors).[32] Bad measurements may also be the result of made-up data (e.g., simply reporting data of good performance that never occurred). If the measurement framework lacks sufficient mechanisms to verify the provenance of the measurements, then the trustworthiness and value of the measurements for decision-making is at risk.[33]

---

[32] For example, measurement systems are not faultless. Sometimes measurements of speed may result in impossible measurements, i.e., speeds that are physically infeasible to have been delivered. Often data cleaning involves excluding outliers that are more likely to reflect measurement system errors rather than real outlier performance. Overly aggressive data aggregation (e.g., weighting) or censoring strategies can exclude from the measurement results outlier results that should have been considered. For example, in tabulating household incomes for the U.S. Census, the quantification of income by category is more granular for lower-income households than higher-income. For example, all households with incomes over $200,000 per year are reported as a single category. That decision makes it infeasible to use this data to analyze the distribution of incomes in the upper tail of the household incomes. Such a decision may be justified in light of the focus of interest on the distribution of incomes at the lower tail of the distribution as being more relevant for many policy decisions that might be based on the U.S. Census. However, for assessing the impact of income taxes, that decision proves problematic. See, for example, Census Reporter (n.d.); or for further discussion of the complexity of estimating income/equity effects, see Fixler et al. (2020).

[33] The greater the incentives to provide bad measurement data (because of the adverse implications of decisions based on good data or because of the ease with which bad measurements may be promulgated), the greater the need for verification tools and methods. Having multiple independent sources and methods for generating measurements helps reduce the threat of 'measurement capture' by bad actors since it provides a way to cross-check data (easing in the detection of misleading data) and thereby helps reduce the likelihood that data manipulation will prove a successful strategy. However, since 'fake measurements' will remain a viable cybersecurity threat and cybersecurity is a war game, there is unlikely to be any conclusive solution to eliminate the risk of intentionally bad measurements. It is sufficient at this stage and herein to focus on the measurement problems that may arise even in the presence of cooperating agents who still may make errors and may still fail to (reasonably) agree on what the best measurement approach is to inform a particular policymaking or decision context. There are several ways to prepare and present bad measurements—introduce noise (intentional or unintentional errors) into measurements that simply confuse and distort. Ideally, everyone should be able to agree on what are bad measurements (errors, fake news, etc.), while recognizing that consensus on what constitute the best measurements or a single set of metrics may not be achievable.

As industry constellations and measurement challenges are becoming more complex, an appropriate measurement ecosystem that is capable of embracing an evolving range of contexts and measurement challenges based on lots of different yet valid perspectives that may vary based on the particular decision-making context under consideration, is required. Different measurement strategies reflect different trade-offs in terms of details, focus, cost, and other aspects. For example, earlier we noted the increased controllability of drive or fixed-probe testing methods at the expense of higher cost testing for large measurement coverage (in terms of space or time), with crowdsourced strategies proving more cost-effective in those contexts.

With broadband, there may be numerous questions of interest that are focused on the narrow question of the data rate. The relevant question may be whether the service consistently delivers at least some minimum level of service or what the average or peak level of the service is. Those measurements may be directed at evaluating the experience of broadband users when they were using the service, or to determine whether the service promised by providers was generally available (so also at times when the broadband user was dormant). The relevance of any deviations in terms of worse or better-than-promised performance may depend on the application being delivered, as well as the (subjective) perception of different end-users or the preferences of application/content providers and broadband service providers. For consumers, the likely focus is on whether the end-to-end experience was what was expected, sufficiently better than next-best-alternatives to justify the choice (price-quality trade-off), etc. Conversely, different intermodal broadband providers may prefer measurement strategies that accentuate the strengths of their technologies.[34]

These explanations illustrate that building an appropriate measurement ecosystem is based on the recognition that multiple valid perspectives are possible and desirable, as well as the fact that there will be (and should be) measurements that are accepted as part of public/policy/market discourse that are legitimate but may differ (e.g., on per-link or end-to-end basis) and not be easy to compare/contrast or aggregate/combine. This is due to several reasons. For example, there are different weightings of measurement factors/dimensions that may be appropriate and once summary statistic is created, it may not be easy to disaggregate or understand implications of different weightings. An algebra that enables to make sense of a multitude of measurements is required.

When considering the implications for specific policy issues, it is worth acknowledging that asymmetric information has always posed a challenge for regulators. The entities that are the focus of regulation, their competitors, many of their customers, and other interested market stakeholders typically possess detailed relevant information that is not readily available to regulators or shared among all participants. The participants may have different perspectives on what they want regulators to do, and hence, different incentives regarding what information to share with regulators. Furthermore, since the future is unknown, regulators

[34] For example, cable providers took advantage of the ability to offer extreme high-speed downstream data rates not feasible on Digital Subscriber Line (DSL) networks. That capability was observable in speed measurements reporting average data rates over initial five-seconds of a connection. DSL providers preferred speed measurements that averaged over longer-time periods.

must always make decisions in the face of uncertainty and incomplete information.[35] Moreover, when regulators are compelled to make decisions on the basis of incomplete, asymmetric information, their decisions are open to challenge from those who may disagree with the decision. Such disagreements may be the basis of misunderstandings rooted in incomplete information (which may be viewed as a problem with transparency) or strategically motivated. Consequently, asymmetric information problems are fundamental to the regulation.

An interesting question is whether the ability to undertake more fine-grained and real-time network management and the fact that this makes it feasible for regulators to require the production of more detailed and fine-grained performance data will make the asymmetric challenges more or less difficult to resolve. Our conclusion is that the answer is, at best, ambiguous. On the one hand, to the extent there is a shift to market-based regulation, the detailed regulatory-control that regulators may be called upon to administer may decrease. More of the detailed setting of investment, pricing, and product design decision-making are left to firms to determine, with regulators shifting their attention to ensuring that market outcomes are supporting public efficiency and equity goals. On the other hand, the opportunity for questions to arise and the potential to collect data to impact the outcome of decision-making regarding whatever the question is will increase also. Expanding the ability to collect more information will certainly change the decision-making environment, but will not necessarily simplify the asymmetric information or decision-making challenges. More attention will be focused on deciding what are the right measurements to undertake and adjudicating among different analyses based on disparate measurement strategies and data.

In the following, we explore how three sets of policy issues may be impacted by the changing measurement ecosystem.

3.1. Universal service and connectivity targets

Broadband is widely recognized as essential for participation in modern societies. This is reflected in Universal Service Objectives (USOs) for broadband and broadband targets motivated by industrial policy objectives. Whereas related public policies and regulations in the EU and the US have acknowledged technology evolution, market developments, and changes in user demand, they render knowledge about the state of broadband deployment and availability a necessity.[36] More specifically, such knowledge is indispensable for (i) the

[35] The future is uncertain in so far what will happen is unknown. That may be a mix of uncertainty amenable to stochastic forecasting among future known potential states, but also unknown future states, the very existence of which may not be known. This point calls to mind the quote from Donald Rumsfeld regarding "*There are known knowns – there are things we know we know….we also know there are known unknowns – that is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.*" (Graham, 2014)

[36] In the EU, the EECC introduced a universal obligation for broadband by installing a basic service (called an "*adequate broadband Internet access service*") capable of supporting a set of services (including email, search engines, education online tools, online newspapers, professional networking, internet banking, eGovernment, social media and instant messaging, call and video calls in SD) listed in Annex V of Directive (EU) 2018/1972. More recently, Article 4 of Decision (EU) 2022/2481, which established the EU's Digital Decade Policy

reliable identification of unserved and underserved areas, (ii) the identification and assessment of cost-efficient deployment and upgrade strategies to achieve the targets, and (iii) to inform decisions related to USOs or other broadband connectivity targets.

On a very high level, there are several challenges associated with the technological definition of coverage and availability. As of today, there is no standard practice across NRAs,[37] but a diversity of KPIs such as coverage scores, coverage experience, and availability. Moreover, as we discussed above, radio coverage is only one part of the story. The growing complexity of mobile networks and the value chain constellations renders availability measurements more complex. When considering the capabilities of measurement approaches, drive and walk testing is infeasible at scale. However, crowdsourcing measurement capabilities are more limited in sparsely populated regions. Figuring out how best to balance measurement strategies and techniques across different market contexts will require further research, but it is likely that a mix of techniques will be needed. For example, limited drive tests and crowdsourced measurement strategies may be combined, and also may be integrated with remote sensing methods (e.g., use of satellite image analyses to facilitate better estimates of crowdsourcing data collection efforts).

There are increasingly more cases in which NRAs are backing up their universal service and other public policies for broadband with network performance measurements using their own mobile apps (e.g., the Federal Communications Commission (FCC) in the US or the Bundesnetzagentur (BNetzA) in Germany). These help them verify the accuracy and reliability of national broadband maps and other annual reports. Other policymakers rely on drive tests (e.g., such as the regional government of Asturias). Box 1 provides a brief overview of different NRA strategies.

Box 1: Examples of Broadband Measurements by Policymakers

In the **U.S.**, the FCC has made available a speed test app (FCC, 2022) that users may use to submit measurements about their WiFi or mobile internet connection to help verify the accuracy and reliability of the FCC's National Broadband Map (FCC, n.d.). As this map constitutes the baseline for the distribution of universal service funds, users who believe that their mobile performance showed on the map is inaccurate may submit Challenge Speed Tests from the app. This incentivizes that interested parties submit up-to-date measurements using the app.

A similar approach is used by the BNetzA in **Germany**. The *Breitband Messung* app enables users to measure mobile performance and help identify *dead spots*. BNetzA uses these results to elaborate their annual reports on the state of mobile broadband (zafaco, n.d.).

In **Spain**, the regional government of Asturias, a mountainous region in the north of the country with many hard-to-serve areas, has recently undertaken a drive test campaign to identify underserved areas (Principado de Asturias, 2022). The results of the campaign are available on the public Geographic Information System

Programme 2030, states: "*all end users at a fixed location are covered by a gigabit network up to the network termination point, and all populated areas are covered by next-generation wireless high-speed networks with performance at least equivalent to that of 5G, in accordance with the principle of technological neutrality*". In the U.S., the Infrastructure Investment and Jobs Act has put forward huge subsidies to deliver high-capacity broadband across the U.S. See, for example, Stocker et al. (2023).

[37] Ofcom, for example, sets a minimum signal threshold of signal strength for voice coverage of −115 dBm in LTE800, −103 dBm for UMTS2100, and −93 dBm for GSM (Ofcom, 2015). Other NRAs may set different thresholds to compute voice coverage.

(GIS) (Principado de Asturias, n.d.), although the targeted stakeholders of the measurement campaign are the MNOs providing services in the region so that they can identify areas for improvement.

At the **EU level**, the Joint Research Center (JRC) of the European Commission analyzed the existing territorial disparities across the EU (Proietti et al., 2022), including the digital divide. For this analysis, they used Ookla crowdsourced data, and found a positive relation between access speed and the degree of urbanization, confirming that best physically connected areas (i.e., with transportation infrastructures) are also the most connected from the digital point of view (Macdonald, 2023).

Identifying unserved and underserved areas by either means faces several challenges, including (i) the definition of availability, (ii) the incomplete picture reported by radio coverage measurements, and (iii) the diversity of KPIs.

Definition of availability. The definition of availability, which is often summarized in terms of coverage is complex as it may vary across different contexts, service providers, and technologies.[38] In LTE networks, one possible definition for coverage is using the Reference Signal Received Power (RSRP), a proxy for received signal strength. As per the LTE standard, the threshold for the minimum measurable RSRP is −140 dBm (Dahlman et al., 2013)—if this threshold is not met, a location is considered not covered.[39] However, such threshold criteria might be insufficient for gaining meaningful insights regarding the QoE for users of different applications. For example, the performance of an application like light web-browsing is less demanding than an application like video-conferencing. Additionally, a stronger received signal will provide better performance during peak periods. Hence, the selection of any such static threshold is problematic, and more so because it would be operator-dependent (Frias et al., 2020).[40]

Going beyond radio frequency measurements. A potential way to solve the problem of defining coverage in terms of RSRP as above is to rely on richer measurements, like throughput or application experience. However, this would come at the cost of increased measurement complexity. Coverage defined in terms of RSRP is a passive measurement that does not involve devices to send data through the networks, as it relies on information broadcasted by base stations (as opposed to throughput). RSRP measurements are easier and quicker to collect at large scale, whereas throughput tests such as speed tests are active measurements that entail more complexity, take longer to measure, and consume more network resources devoted to collecting the measurements. Opting for such more complex

[38] Even the concepts of availability and coverage are exchanged in KPIs reports. For example, availability may be defined in terms of percentage of time a service is operating and capable of supporting a particular service. Availability may also be defined in terms of coverage (i.e., probability or percentage of locations in an area where service is available). Coverage may be metric used to define availability, as for example, percentage of locations where the received signal strength exceeds some threshold. For further discussion of KPIs for coverage or availability in mobile networks, see Casas et al. (2015) or Krasniqi et al. (2019), which illustrate the multitude of KPIs that are under consideration.

[39] For reference, the mode of the measured distribution of RSRP of most operators is around −110 dBm (Frias et al., 2020).

[40] Note that operators with more spectrum have larger cell sizes and hence lower RSRP reported values. Operators with comparably poorer spectrum portfolios need to compensate the reduced bandwidth by increasing the cell density, leading to comparatively smaller cell sizes, and larger RSRP reported values as end user are closer to the antennas. As this shows, it is not trivial to find meaningful threshold values in a world where deployment scenarios and network topologies differ.

measurements to define availability in terms that may be more meaningful for assessing the QoE would likely come at the expense of the spatio-temporal resolution of the measurement program. Moreover, in contrast to measurements such as RSRP, which are standardized in 3GPP protocols, there is no single industry-consensus standard method for measuring throughput. Therefore, multiple different measurements may be required to define network availability for specific services and applications.

KPIs diversity. As there is no unique way to measure network availability, different stakeholders estimate it in different ways. Companies conducting crowdsourced measurements often use the term *availability* to refer to the temporal availability of a service (i.e., its uptime). The spatial availability of the service is more typically referred to as reach or coverage. For example, as summarized in Table 1 in the Appendix, Ookla defines the Coverage Score (1-1000) based on the operator footprint (understood as the percentage of locations where an MNO reports offering coverage) and service (the probability of access to 4G services). OpenSignal's KPIs for this are Reach (1-10), which is the proportion of locations where users were connected to 3G/4G/5G networks relative to all the locations visited by the user (in a given time frame), and Coverage Experience (0-10), which is defined as the proportions of MNOs' locations with 4G/5G coverage relative to the locations covered by any MNO. Although OpenSignal's Coverage Experience indicator and Ookla's Footprint indicator intend to capture similar realities, they are unlikely to be equal, or even, highly comparable metrics, as there are many ways to compute either of them.[41] Moreover, how the ideas underlying the definition of the metric translate into specific KPIs in a range, for example, 1-1000 or 0-10, is undisclosed by the companies, which complicates the understanding and the assessment of the comparability and compatibility of the different metrics.

In addition to the challenges mentioned above, there are specific challenges associated with each measurement method in the context of the policy issue under consideration.

Drive testing scales poorly. As described in Section 2, drive testing requires dedicated personnel, specialized equipment, and significant time and effort to conduct tests by physically driving vehicles equipped with measurement tools. This resource-intensive nature of drive testing poses challenges in terms of cost, manpower, and logistical requirements, particularly when trying to cover large geographic areas.

Mobile crowdsourcing needs further research. MCS has limitations when it comes to estimating unserved areas. Crowdsourced measurements are collected from user devices, resulting in a higher concentration of measurements in densely populated or frequented areas such as cities. Hence, in rural areas, the data collection capacity through crowdsourcing is more limited. Therefore, problems to differentiate between so-called deadzones (i.e., areas without coverage) from areas where no data has been reported arise. For example, if there is a temporary lack of mobile signal within an area, MCS approaches that collect data in the background will identify such an area as a deadzone. This may lead to false insights but can be mitigated once the connectivity has been established again. In contrast, those areas with no registered measurement attempts would still present a problem as they would be identified

[41] For example, in the way the locations are defined to compute the proportion of locations afterwards.

as deadzones, which may or may not be true. This points to the fact that a lack of registered measurements attempts is necessary for an area to be identified as a deadzone, but it is not sufficient.

3.2. Consumer information asymmetries

With the increasing reliance on mobile connectivity for various social and economic activities, including work, communication, and entertainment, mobile broadband measurements play an increasingly vital role in enabling consumers to verify that they receive the experience they desire based on the promised quality and performance from their service providers. By conducting measurements, consumers can assess important network performance parameters such as DL/UL data rates and latency, and compare them against the advertised claims of their service providers. Furthermore, understanding differences between services offered by various providers is essential for consumers to make informed decisions. Mobile broadband measurements provide objective data insights that can help consumers compare the performance and reliability of different networks or service plans by different providers as well as the coverage each service provider offers in their area.

To this aim, NRAs may mandate service providers to disclose and publish complete, comparable, and reliable information on the offered QoS[42] according to specific measurement and quantification approaches. To benchmark the performance of different service providers, mandating the use of the same methodology is essential for the measurements. Typically, the information is presented in terms of 'mobile coverage' and DL and UL data rates, as these indicators seem to be easy for consumers to understand although other KPIs also have a large impact on the end-users' customer experience (i.e., QoE), such as latency, jitter, and packet loss.

In the EU, the legal framework that has, among other things, guided QoS has been reviewed to harmonize what is measured and how. Adopted in 2018, the European Electronic Communications Code (EECC) specifies parameters that apply to mobile network services.[43] Following the EECC, transparency obligations mandating QoS measurements are under review in many Member States. Relevant measurements may be conducted with different methods, and there are examples of use of all three approaches we described in Section 2 (unattended probes, drive tests, and crowdsourcing). For example, in Spain, the transparency obligations (MINECO, n.d.) require MNOs to deploy unattended probes to monitor their network performance and provide quarterly national aggregate statistics of 95, 50, and 5 percentiles of DL and UL data rates on their websites.[44] In France, the national NRA ARCEP

[42] For example, Recital 271 of the EECC (Directive (EU) 2018/1972) states (at p. 87): *"Where the quality of services of publicly available interpersonal communication services depends on any external factors, such as control of signal transmission or network connectivity, national regulatory authorities in coordination with other competent authorities should be able to require providers of such services to inform their consumers accordingly."*

[43] See Annex X at pp. 203-204 in the EECC (Directive (EU) 2018/1972).

[44] See, for example, MÁSMÓVIL(2023); Movistar (2023); Orange (2023); or Vodafone (2023).

undertakes its own measurement campaign with drive tests[45] and publishes the results on the website *Mon Reseau Mobile* (Arcep, n.d.), which displays a map with the measurements conducted and a separate dashboard with aggregate statistics. To provide a non-EU example, the Perouvian NRA Osiptel uses crowdsourced measurements for the website *Checa tu señal*[46] that facilitates spatially disaggregated information on network performance.

However, the use of mobile measurements to help consumers make more informed decisions gives rise to several challenges. In the following, we explain three major challenges.

More granular data and customized/location-specific insights are needed. Since people live and work in specific areas, national aggregates may not serve as good proxies for the performance in these specific areas. Therefore, they may provide little guidance to consumers for their location-dependent decision-making. National aggregates may provide relevant information for some—e.g., those often traveling across the country—but not relative to the places that users visit more frequently. Although some NRAs are making efforts to provide consumers with more detailed information about mobile network performance[47], and these efforts deserve recognition, the geographic scope of the available data is still insufficient for end users to fully understand the specific service quality they can expect in a particular location. For example, the maps may indicate that an MNO offers service at some locations within a geographic area, and the geographic areas may vary in size. That does not allow a user to understand whether the MNO offers acceptable service at the home or other precise locations where the user may expect to use the service. Indeed, as users of cellphones are well aware from personal experience, service quality can vary significantly even over distances of a few feet and be impacted by whether a user is indoors or outdoors, the time of day, and myriad other physical and network-related factors that may impact the quality of the radio-channel or higher-layer protocol performance. To determine whether an MNO will meet the QoE requirements of any given consumer at any particular location, more granular data is needed.

The ability to customize network performance information is highly dependent on the measurement method used. Different measurement methods provide varying levels of customization and detail in the collected data. Drive testing generally offers a limited level of customization as it fails to capture the temporal profile and dynamics of network performance. On the contrary, unattended probes may provide high temporal resolution but their capability to provide spatial resolution depends on the density of deployed probe equipment.[48] MNO benchmarking has been one of the first use cases of MCS measurements

[45] ARCEP is also exploring the use of crowdsourcing methodologies and displays results based on crowdsourced measurements from external sources.. In their methodology documentation on QoS, ARCEP states that these results need to be interpreted differently from those of ARCEP’s measurement campaigns as crowdsourced measurements show large variations depending on many parameters. (Arcep, 2022).

[46] Osiptel, the Perouvian regulator, provides a website (https://serviciosweb.osiptel.gob.pe/CoberturaMovil/) that allows users to check the service coverage for different operators over different geographic locations (Osiptel, n.d.).

[47] For example, ARCEP disaggregates the data into touristic, dense, intermediate, and rural areas.

[48] For example, the Spanish regulation requires the number of probes to be proportional to the population of the region, resulting for example in only a few probes deployed in large regions such as Andalucía with over 8.5 million population and over 87,500 km^2.

due to their high spatio-temporal resolution. One of the MCS pioneers, OpenSignal, utilizes mobile crowdsourcing to gather data from end-users and provide detailed insights into network performance across various locations. By harnessing user-contributed data, OpenSignal helps fill the void in understanding network performance at a more granular level and developing higher-resolution spatial statistics that can benchmark MNOs' performance for any specific location.

More detailed performance data are required for consumers to make informed decisions. Unlike in fixed networks, in mobile networks consumer choice is not primarily driven by the advertised speed/performance. As mobile networks are more subject to performance variations, MNOs typically make their marketing efforts on promoting the access technologies available, such as 4G or 5G, rather than specific speed or performance claims. To assist users in verifying their access to the promised technologies, mobile phones typically display the best available access technology through an icon positioned at the top of the screen. This visual indicator allows users to understand when and where these promised technologies are available.

Whereas this suggests an easy and helpful way to indicate the quality of the services offered, the representation of radio access technologies on mobile devices is often simplified with the mobile generation they belong to.[49] These simplifications help streamline the display of access technologies on mobile devices, but they can also lead to confusion due to the variety and scope of different access technologies, deployment scenarios, and capabilities within a given generation. 5G, as an evolutionary technology building upon previous generations, has increased the complexity of this problem, in particular, since Non-Standalone (NSA) 5G upgrades the RAN but uses the LTE core network. Therefore, in NSA deployments, smartphones may not always detect the availability of 5G services due to the exchange of information occurring through the LTE control plane. To address this, the 3GPP has standardized a 1-bit indication for broadcasting 5G presence in a cell, but it can be set independently of the actual functionality provided by the network, leading to potential mismatches between the (potentially misleading) indication and full 5G capabilities. Some operators' controversial use of *5GE* for enhanced 4G technologies has raised concerns about misleading users, and has incentivized the use of other icons (e.g., 5G+, 5GUC, 5GWB, etc.) to reflect the increased capacity when mid-band or mmW spectrum is available.[50] This also highlights the need for clear and transparent communication of actual network performance via mobile performance measurements to avoid misleading or confusing consumers.

Insights on network performance need to be presented in an understandable way for consumers to be capable of making informed decisions. Most end-users lack the capability or technical knowledge to make and interpret measurements and measurement results themselves, without interpretative assistance. End-users may thus rely on measurements conducted by third parties. As we explained in previous sections, third-party stakeholders such as OpenSignal, Netflix, or other edge application providers embed measurement

[49] For instance, the Universal Mobile Telecommunications System (UMTS), encompassing the initial releases of 3G, is commonly displayed as 3G, whereas enhanced 3G technologies, like High-Speed Packet Access (HSPA) are represented in some countries as H.

[50] For examples in the US, see, for example, Johnson (2022).

capabilities within their applications and provide insights on network performance. Governments may provide end user performance apps and publish aggregated summary information with accompanying interpretative analyses. Moreover, MNOs may publish insights into their networks, characterizing how their services have performed, and often accentuating their successes in terms of growth and customer satisfaction reports. With the multitude of information sources available, each providing different KPIs and metrics, and the increasingly complex networking and service landscapes, consumers confront a complex decision-making environment. Figuring out what is the right information or even what is the right question to ask to allow a consumer to knowledgably evaluate a consumer's option is difficult. Choosing among one or two options may offer too little scope for choice, but choosing among thousands of options that are difficult to understand and compare does not make consumer decision-making any simpler or better (see Schwartz, 2015).

To address this issue, there is a need for enhanced analytic capabilities within the measurement ecosystem. These capabilities would aim to render the wealth of information more easily comparable and facilitate assessment of its provenance. Third-party sources should be able to synthesize and summarize the data in ways that makes it easier to understand for consumers, and since different consumers have different preferences, there ought to be multiple such sources. To facilitate better decision-making there ought to be well-documented suites of standardized KPIs that can be evaluated and presented in readily comparable ways, utilizing visual representations, and facilitating network comparisons.[51]

3.3. Other Challenges: Value Chain and Resource Sharing Dynamics and Concerns Regarding Privacy and Security

The state and evolution of mobile technologies has not only changed and restructured the nature of the measurement challenge, it has also rendered value chains more complex, diverse, and dynamically changing (see Section 2.2). Key technology changes including the growing importance of small cells, AI/software control, and on-demand provisioning have changed the measurement challenges that emerge from various policy challenges related to market performance oversight, obligation enforcement, and dispute resolution.

As (more) intelligence is embedded along the value chain and at multiple levels, more dynamic, flexible, and fine-grained (local in every dimension) customization and sharing of resources (owned and managed by a more diverse set of actors) becomes possible. On the one hand, this is reflected in growing potential for new business models and hence for customers to purchase almost Anything-as-a-Service (XaaS), implying that unbundled and

[51] Whereas aspects of digital literacy are important regarding participation in the digital economy and are subject to discussion in a variety of contexts, similar discussions emerged in the context of Internet privacy and the way to display related aspects in a clear and understandable fashion to the 'average user'. Privacy researchers have directed significant attention to ensuring that processes and capabilities implemented by businesses to protect user privacy on-line be effective, which includes being usable (i.e., understandable, easy to employ, etc.). Many firms are tracking the effectiveness of their efforts using a variety of metrics (see Tene & Culnan, 2022), but Cranor and Habib (2023) call for more to be done.

bundled business models are simultaneously competing.[52] On the other hand, as resource sharing across different actors becomes more important and diverse across contexts, more markets are in need of more sophisticated technical and economic performance measurement data to efficiently manage and effectively monitor the resource sharing. Expanded options for what to self-provision versus what to outsource means an expanded array of alternatives to model and evaluate in terms of pricing, strategic implications, and other effects. These developments underscore that evidence-based decision-making processes will be based on performance measurements collected and interpreted by multiple stakeholders representing diverse perspectives and based on diverse measurement strategies and methods.

The industry structure itself has become increasingly complex and fluid. The distinction between firms and markets is no longer as clear-cut, with the dynamic evolution of companies and industry structures. The traditional vertical and horizontal categorizations do not adequately capture the intricacies of the mobile broadband ecosystem, which involves multiple stakeholders, collaborations, and ongoing changes. Beyond market and technology changes, we anticipate policy changes that give rise to novel measurement challenges. For example, spectrum reform may make new bands (mmW to terahertz) available for commercial use and potentially more and new ways of sharing across regimes and services.[53] Other (potential) reforms may be related to network sharing and interconnection policy, implying that policymakers (may) embrace new active sharing models. Moreover, (potential) new regulations regarding network/traffic management (e.g., in the context of network neutrality and 5G+-based network slicing; see, e.g., Stocker et al., 2020) or novel forms of interconnection rules (e.g., mandatory edge provider payments to ISPs being considered in EU; see, e.g., Stocker & Lehr, 2022) may be deemed needed to address concerns over potential abuses of market power and related concerns. Any such reforms would give rise to new monitoring obligations with attendant transparency/disclosure rules.

In view of the above, we identify two sets of challenges regarding the role of measurements.

Granular measurements from potentially an expanded array of stakeholders and an algebra to make sense of them are needed for effective market oversight, obligation enforcement, and dispute resolution. From a policymaker's perspective, the need for more granular and real-time measurements from several vantage points (and actors) is growing. Similarly, information sharing of collected insights across the value chain are necessary. Finally, to make sense of these insights, developing an algebra will be indispensable. Market performance oversight (e.g., dynamic and active infrastructure sharing, network resource

[52] Historic taxonomies of cloud services differentiated between Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Sofware-as-a-Service (SaaS) offerings. IaaS involved selling basic components such as transmission or router capacity needed for constructing more complex networked services; PaaS involved selling higher-level services that could be used to support customer applications such as private clouds; and SaaS involved providing cloud hosted-software application services. Further developments tailored the offerings to address the specialized needs for additional functionality or for specific market segments such that almost any business resource or bundle of resources may be purchased as a service, instead of self-provisioned. Hence, the emergence of XaaS. See, for example, Duò (2023) or Duan et al. (2015).

[53] Spectrum reform may aim at addressing sharing with sensing apps (radio-astronomy) and expanded Non Terrestrial Network (NTN) options, while seeking to liberalize legacy frameworks to enable more and new ways of sharing across regimes and services.

allocation), obligation enforcement (e.g., coverage obligations in spectrum licenses, spectrum use) and dispute resolution (e.g., SLAs, interconnection tussles, net neutrality) become newly more important in view of expanded options for more dynamic and active infrastructure sharing and network resource allocation (across different competing actors). Supporting these in active markets will only be feasible efficiently if an appropriate measurement ecosystem is put in place. This ecosystem should enable the assessment of all actors involved in collectively delivering end-to-end services (and thus to the customer experience), ensuring they meet their contracted service obligations. That ecosystem has to make it viable to differentiate between false and truthful claims, and hence, designed so that all of those involved are appropriately incentivized to take relevant measurements and disclose relevant insights. In the strategic context in which multiple participants will be engaged in co-opetition (i.e., businesses that are rivals yet must work together) to support network services, it should be feasible for all legitimate businesses to agree on what are misleading and incorrect measurement strategies and performance data, while still failing to agree on what are the best measurement strategies.[54]

Privacy and security considerations must be carefully considered. Relating to data collection and disclosure, advancing monitoring and measurement capabilities give rise to trade-offs related to concerns regarding lawful data access, and with this to security and privacy. With expanded, finer-grained information, the surveillance society/security risk issues of making data public for everyone loom larger, as does the potential for abuse by strategic sampling, tampering with (potentially generating spurious measurement data), or misrepresenting (bad analysis) of data for strategic purposes. While it is hard enough figuring out the right answer in complex situations and constellations, that problem can become harder if there are efforts underway to distort/bias the effort to identify the right answer.

Additional granularity and complex measurements will also raise concerns in what should be shared with whom. Operators and end-users will have valid privacy and security concerns that access to fine-grained data be restricted (i.e., not fully open). For example, restricting access is necessary to protect Personally Identifiable Information (PII) and against 'surveillance economy'.

In addition, it is necessary to avoid critical infrastructure from being targeted by attackers or strategically sensitive information underlying competitive advantage (e.g., SDK proprietary code) be disclosed. The latter would harm competition incentives if no ability to have proprietary information. On the other hand, it may provide the basis for excess investment or market power if intellectual property is not shared at all.

Data governance will require a tiered structure because (i) without specialized expertise, raw data/measurements are too complex to interpret, hard to manipulate, etc. (hence, expect AI and other software tools such as dashboards and UIs to make them accessible); (ii) not everyone needs to (or wants to) know everything, this would raise safety and security issues (note that different stakeholders have different information needs depending on their (specific) decision contexts); (iii) the costs of measurement infrastructure should be minimized. Whereas these costs should be shared as much as possible, sufficient

[54] For a discussion of co-opetition, see Brandenburger & Nalebuff (1996).

independence to allow cross-validation and disentangling any potential strategic abuses of data (e.g., fake news created from good data by selective reporting or from bad data developed via selective sampling) needs to be preserved.

4. Conclusion & Outlook

As we are heading towards a much richer world of potential measurements, there are also accentuated, or more intense, asymmetric information problems. These, in turn, emphasize the role of measurements for regulation and optimal decision-making. Determining "who knows what," "what is knowable by whom," or "who needs to know what" all become more challenging questions in the more complex future. Some measurements might be easier for service providers to make, while other measurements might be easier for edge providers. Some other measurements might only be feasible if made by end-users (or at least under the control of end-users). And, the ease/cost of undertaking the measurements may not naturally align with the incentives or needs of those seeking insights that the measurements might inform.

In short, asymmetric information challenges emerge as a regulatory design and complexity challenge. These challenges are aggravated by the growing importance of incomplete information because the unknown future matters more. In complex systems in which the number of future states is vastly greater, the transition times between states are shorter, and there is greater heterogeneity in risk tolerance (e.g., regarding potential benefits versus costs, and irreversibilities), it may be either more or less necessary to act with incomplete information—depending on the context. Not every measurement that is possible should be undertaken and not all information that is known by some should be shared with all. Total measurement (even if economically feasible) would enable a regime of maximum surveillance, while total public disclosure of all measurements would pose an unacceptable threat to privacy and security. Measurement design decisions need to consider the incentives of the parties to acquire and *appropriately* share costly information.

In this paper, we focused on mobile broadband services. We explored the need for and the necessity of enabling a third-party augmented measurement ecosystem that is needed to complement and supplement government and service provider measurements. A central characteristic of such an ecosystem is its flexibility, particularly in involving all key actors in the ecosystem—governments (e.g., NRAs), service providers (e.g., MNOs), and third-parties (e.g., end-users or edge providers of content and applications)—to facilitate context-aware measurement strategies as well as required data sharing and analytics. Depending on the question to be asked and the context (e.g., regarding application, location, or time), appropriate measurement strategies and data sharing requirements across different entities may vary.

Our paper has laid out that there is a need for a more complex, multi-party (multi-source) broadband measurement ecosystem with more granular data to allow different actors to investigate more complex and nuanced decisions (from "can you make a telephone call anywhere?" to "is service provider A better than service provider B for application Y at location Z now?”). Along with more granular data, we will also need better ways to assess the provenance and understand the measurement methods used to collect, summarize, and analyze the data. This will be needed to evaluate the more complex and multi-party agreements for sharing network resources for their consistency with policy goals to promote universal service, good QoE, and competition (goals which are all important, but involve trade-offs that may be difficult to resolve).[55]

In the future, many more measurements will be needed for many more services and contexts to coordinate the investments, behavior, and experiences of many more stakeholders—e.g., end-users getting what they need and what they paid for; ISPs and other businesses that need to coordinate operations to deliver an end-to-end service efficiently at minimum cost and suitable quality; and government regulators seeking to enforce public policies and regulations. Whereas all actors need information, individual actors will be able to anticipate the needs of other actors and that will factor into their ex ante incentives to cooperate in evolving the measurement ecosystem.

When focusing on the (future) role of different stakeholders, we believe that governments should be limited in what they should be expected to be able to provide or mandate disclosure of—this is due to various reasons including costs, privacy, and other policy reasons. The goal is to ensure well-functioning markets based on market-based regulatory policy, not command-and-control-style heavy-handed regulation. In fact, they should carefully assess data sharing obligations and mandates for information disclosure.

Service providers may be in the prime position to make valuable measurements. They have lots of measurement capabilities, potentially expanding in scale and scope, which they should not be expected to deploy. This implies that lots of possible measurements will not get made because there is no compelling need and associated cost would be prohibitive (measurements are expensive). On top of that, there are several reasons related to security, privacy, or strategic reasons why these providers should not be compelled to disclose some of their measurement results. In fact, mandating such disclosure would constitute a violation of sound market-based regulatory policy.

Going forward, third-parties are slated to play a more critical role in the measurement ecosystem. However, since most individuals or small businesses lack the capability and

[55] There may be multiple parties with conflicting yet valid perspectives. For example, a monopoly may lower total static costs and be better able to deliver on universal service goals quickly; but competition may lower dynamic costs (even at the expense of excess duplication of some investments) and be more robust for addressing future uncertainty. Where should facilities-based competition (duplication of capacity investments) occur and where should sharing occur? And which parties should be involved in controlling those shared assets and be responsible for the costs of providing those assets? The evolution of CDNs provides a useful illustration of how those costs have shifted and the needs for cross-layer optimization in mobile networks that can lead to different traffic management practices between service providers being pro-competitive (i.e., allowing intermodal competition between MNOs with slightly different network architectures being capable of offering products that deliver a QoE that makes them reasonable substitutes to consumers so that price competition between those MNOs is intensified). If such flexibility in traffic management is blocked by network neutrality regulations (and which is more important in wireless where RF is likely to be a scarce resource), then competition and innovation may be harmed by a regulatory intervention that was intended to safeguard competition.

knowledge to make and interpret measurements themselves, it is to be expected that all end-user measurement at some level will likely involve either governments, MNOs, or other third-parties like OpenSignal or Ookla, or other stakeholders (e.g., Netflix or other edge providers of content and applications embedding measurement capability in their apps). What is particularly interesting is to what extent third-parties enable not only context-specific (e.g., application-aware or application-specific) but also user-friendly measurements, reflecting a specific information need.[56]

Taking this into account, from a high-level perspective, our paper yields the following three key insights.

First, the measurement ecosystem has become more capable and complex. There is a need to recognize that the supply of mobile measurements has evolved. The increased capabilities (i.e., ability to collect more measurements and use those measurements for more fine-grained information acquisition and real-time decision-making) is both enabled and necessitated by the expanding capabilities of our digital infrastructures. The plural is intentional here. The digital infrastructures we need for the digital economy will need to provide more than just connectivity. It will also have to provide networked access to the computing and storage resources to enable information technologies to augment (automate) a growing number of economic activities in businesses and society. Enabling these digital capabilities or tools alone does not ensure that those will be used to benefit society, and so regulatory policies will be needed to identify and intervene when tools are missing or available tools are being misused.

Second, there is no single best measurement source, nor should there be. Multiple sources, reflecting multiple divergent perspectives/vantage points are needed for good evidence-based policymaking in the multistakeholder digital future that we hope to see evolve. In that world, there will be an increased need for a larger number of industry players to share resources and coordinate their actions, while also ensuring healthy competition along and across the value chain, to deliver the high-quality end-to-end digital experience desired and needed. If competition is robust, innovation continuous, and network architectures adapt as expected, different detailed data and measurements will be feasible and needed to address different policy questions. In that data rich/information rich environment the potential for (new) asymmetric information problems (e.g., competition failures, fake news problems, inefficient risk management strategies, etc.) will be significant. This means that mobile broadband measurement is becoming (and unavoidably so) more important and strategic.

Third, measurements are used strategically. By strategic, we mean that there are expanded opportunities and incentives to potentially distort the measurement ecosystem at the level of measurement collection, processing and aggregation, reporting, analysis, and decision-making (based on the data).[57] With lots more data collection opportunities and lots more data

---

[56] For example, OpenSignal focuses on providing user-friendly mobile measurements that are more specific to the usage context (and measurements need context).

[57] In computer science, "GIGO" is shorthand for "garbage in, garbage out" to informally explain how bad inputs can result in bad outputs. With a more complicated measurement value chain with more steps between what is

being collected, ensuring measurement costs (overhead) are managed appropriately (overhead is not excessive) will be a challenge. Figuring out how to incentivize and share the costs of measurement overhead will represent its own policy challenge, as well as how best to manage the data to protect privacy and security. With lots more valid and distinct ways to obtain measurements about similar issues (e.g., active and passive traffic probes at different locations, remote sensing, and alternative direct and indirect measurement options), the potential for measurement errors and measurement incompatibilities will expand.[58] Furthermore, the capabilities to collect and interpret measurements will not be symmetrically distributed across industry participants. The MNOs will have the best (position to gather) information about what is going on in their networks, at least in principle, but there is much that even they do not know,[59] and what is going on in their networks does in many cases not provide sufficient information for evaluating end-to-end performance.[60] Moreover, in practice, some third-party service providers that offer services to MNOs or on-top of MNOs like CDN providers may actually have better information than will end-users or policymakers. Which stakeholders know the most about market-relevant data depends on what market one is focused on and what data is of relevance. Individual consumers know their willingness-to-pay for services, but whether digital platform or edge providers of applications and content services, or third-party market research firms or specialized measurement providers (like Nielsen, Ookla, or OpenSignal[61]) have the best data for answering particular questions depends on the question. In light of the commodification of many network and computing resources, control of real-time measurement data and market intelligence is recognized as a critical strategic asset. At the same time, this emphasizes the need to cultivate and foster a third-party measurement ecosystem to supplement measurements by governments (mostly NRAs) and service providers (e.g., MNOs).

---

measured and how the measurements are ultimately made use of, there are more places along the way to introduce measurement, sampling, or other sorts of data errors or biases.

[58] For example, forensic analysis of passively collected data may provide a viable alternative to active measurements for many questions such as the average per-subscriber use if behavior is not changing. Alternatively, the incidence of congestion or latency impairments may be inferred using different strategies such as round-trip delay or packet-loss rates, depending on the context.

[59] For example, many MNOs rely on specialized measurement service providers or consultants for the collection and analysis of measurement data related to such issues as what applications users may be using, the RF environment (e.g., actual as opposed to modeled propagation and RF utilization characteristics), and other specialized measures. Often those specialized measurement providers rely on proprietary software and the access to measurement details may depend on what services the MNOs have contracted for which may leave MNOs with significant measurement uncertainty. Additionally, the software systems that collect measurements for MNOs are often not well-integrated meaning that just because an MNO could in principle aggregate measurement data it possesses (e.g., related to the traffic behavior at different points in their network and billing data), the MNO may not be able to integrate that data in practice. Note that capabilities to collect and interpret measurements will not be symmetrically distributed across MNOs.

[60] For example, when the end-to-end quality of service of an application suffers due to congestion, the congestion could be arising in the end user's home network (e.g., in-home WiFi network), the MNO's network, or the application provider's network (which may control how the content is delivered to the MNO's subscriber). See our discussion on complex and dynamically changing value chains as well as, for example, Möller and Raake (2014) or Stocker and Whalley (2018).

[61] There are a large number of third-party providers of mobile performance data. Some well-known ones include Nielsen (which has long been known for its media market research services, see The Nielsen Company (n.d.), Ookla (which is now part of the ZiffDavis media company, see Ookla (n.d.-a), or OpenSignal (which is one of the many providers of smartphone apps for measuring mobile performance, see OpenSignal (2021).

As value chains are increasingly complex and rapidly evolving, either legacy or new replacement heavy-handed rules are inappropriate and inadequate. Moreover, isolating regulated from unregulated components becomes increasingly complicated. Despite these complexities, the need to regulate competition, address market harms, inefficiencies, and gaps persists. Information is and always has been a crucial facilitator of markets.

When considering the role of regulation that designs disclosure and transparency (reporting) requirements and measurement infrastructure (e.g., what? Who pays for what? Who controls? etc.), it is obvious that this involves ex ante assessments of alternative designs. The right design will depend on the specific context and different contexts may recommend emphasizing one type of measurement strategy over another.[62] We should expect there to be a collection of policies and initiatives promoting healthy growth of different measurement capabilities ex ante. Indeed, the fact that there will be an expanded array of vested interests in making measurements or ensuring their perspective is part of the discussion will (hopefully) ensure that the measurements used to make decisions (by end-users, by service providers, and by regulators/policymakers) will involve multiple valid measurements (that may not agree).

If questions arise, it will be necessary to use the measurement data at hand and that will require evaluating data from alternative measurement sources that may result in data from measurement A being preferred over measurement B—even if ex ante, the methods used for measurement A were better suited to the problem. For example, drive test data and crowdsourced data are unlikely to be perfectly matched in time and location but both may be useful for a problem and making use of them will depend on complex judgement. All data reveals (some) information, but its relevance in a particular decision-making context may require significant technical expertise to expose and assess. Lastly, measurement is much more than just a decision-theoretic challenge (in the sense that operations research/big data statistical/stochastic optimization academic research addresses it). When economics is introduced, it is a game theoretic challenge wherein you have duelling decision-scientist/engineers with strategic goals that may not align.

The regulatory challenge thus is to encourage the right types of measurements to be made. It will be necessary to have the capabilities and processes to appropriately interpret and make use of the available data, which is unlikely ever to be optimal or what might have been desired if the need had been perfectly anticipated in advance.

The current upheaval caused in the EU by the Digital Decade policy programme renders the EU a great test case. As the EU has a strong desire to forge its single market as a microcosm of global challenge, there is a huge empirical research opportunity with a mix of distinct market, policy, and research efforts across EU countries. Developing appropriate measurement and (cost-)effective measurement ecosystems will be key.

[62] For example, for some questions crowdsourced data from end-user applications is best, for other questions fixed-node testing design and standardized reporting is best, etc.

Ultimately, attention needs to be paid to building a healthy measurement ecosystem—for mobile broadband certainly, but this problem is general to the entire transition to a digital economy and for enabling human control of our increasingly ICT-augmented society/economy. As we will increasingly rely on AI-automated processing of the measurement data (e.g., to determine what measurements to make and how to interpret and present those to human decision-makers; and increasingly, to make automated decisions based on the measurements), we need to be attentive to the risks, costs, and health of that measurement ecosystem at multiple levels. That is the higher research challenge and one that requires multidisciplinary engagement/research.

References

1. Arcep. (n.d.). *Mon Réseau Mobile* [Map]. Autorité de Régulation des Communications Électroniques, des Postes et de la Distribution de la Presse (Arcep). Retrieved August 9, 2023, from https://monreseaumobile.arcep.fr/
2. Arcep. (2022). Questionner la qualité de service mobile avec les mesures crowdsourcées. Autorité de Régulation des Communications Électroniques, des Postes et de la Distribution de la Presse (Arcep). https://www.arcep.fr/fileadmin/user_upload/grands_dossiers/qualite-services-mobiles/questionner-qualite-services-mobile-mesures-crowdsourcees-fev2022.pdf
3. Bauer, S., Clark, D. D., & Lehr, W. (2010). *Understanding Broadband Speed Measurements*. TPRC38: Telecommunications Policy Research Conference, Fairfax, VA. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1988332
4. Bauer, S., Clark, D., & Lehr, W. (2011). PowerBoost. *Proceedings of the 2nd ACM SIGCOMM Workshop on Home Networks*, 7–12. https://doi.org/10.1145/2018567.2018570
5. Bauer, S., Clark, D., & Lehr, W. (2013). Broadband Microfoundations: The Need for Traffic Data. In R. D. Taylor & A. M. Schejter (Eds.), *Beyond Broadband Access: Developing Data-Based Information Policy Strategies*. Fordham University Press. https://doi.org/10.2307/j.ctt13x0851
6. Bauer, S., Clark, D., & Lehr, W. (2020). Gigabit broadband measurement workshop report. *ACM SIGCOMM Computer Communication Review*, *50*(1), 60–65. https://doi.org/10.1145/3390251.3390259
7. Bauer, S., & Lehr, W. (2018). *Measuring Mobile Broadband Performance*. TPRC46: Telecommunications Policy Research Conference, Washington, DC. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3138610
8. Bauer, S., Lehr, W., & Mou, M. (2016). *Improving the Measurement and Analysis of Gigabit Broadband Networks*. TPRC44: Telecommunications Policy Research Conference, Fairfax, VA. https://doi.org/10.2139/ssrn.2757050
9. Casas, P., Sackl, A., Schatz, R., Janowski, L., Turk, J. & Irmer, R. (2015). On the quest for new KPIs in mobile networks: The impact of throughput fluctuations on QoE. In 2015 IEEE International Conference on Communication Workshop (ICCW) (pp. 1705-1710). IEEE

10. Census Reporter. (n.d.). *Table B19001: Household Income* [dataset]. https://censusreporter.org/tables/B19001/
11. Brandenburger, A., & Nalebuff, B. (1996). *Co-opetition*. Doubleday.
12. Clark, D., Bauer, S., Lehr, W., Claffy, K., Dhamdhere, A., Huffaker, B., & Luckie, M. (2014). Measurement and Analysis of Internet Interconnection and Congestion. *Journal of Information Policy*, *4*.
13. Connelly, B. (2021, May 12). How Ookla Ensures Accurate, Reliable Data: A Guide to Our Metrics and Methodology (Updated for 2021). *Ookla Insights Articles*. https://www.ookla.com/articles/how-ookla-ensures-accurate-reliable-data-2021
14. Cranor, L. F., & Habib, H. (2023). Metrics for Success: Why and How to Evaluate Privacy Choice Usability. *Communications of the ACM*, *66*(3), 35–37. https://doi.org/10.1145/3581764
15. Dahlman, E., Parkvall, S., & Sköld, J. (2013). *4G: LTE/LTE-Advanced for Mobile Broadband*. Elsevier Academic Press.
16. Duan, Y., Fu, G., Zhou, N., Sun, X., Narendra, N. C., & Hu, B. (2015). Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends. *2015 IEEE 8th International Conference on Cloud Computing*, 621–628. https://doi.org/10.1109/CLOUD.2015.88
17. Duò, M. (2023, May 15). XaaS: The Anything as a Service Model (With 10 Example Categories). *Kinsta Blog*. https://kinsta.com/blog/xaas/
18. ETSI. (2007). *Speech Processing, Transmission and Quality Aspects (STQ); Estimating Speech Quality per Call* (ETSI TR 102 506 V1.2.1). European Telecommunications Standards Institute (ETSI). https://www.etsi.org/deliver/etsi_tr/102500_102599/102506/01.02.01_60/tr_102506v010201p.pdf
19. Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (Recast)Text with EEA relevance, L 321/36 Official Journal of the European Union (2018). https://eur-lex.europa.eu/eli/dir/2018/1972/oj
20. Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030 (Text with EEA relevance), L 323/4 Official Journal of the European Union (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022D2481
21. FCC. (n.d.). *FCC National Broadband Map* [Map]. Federal Communications Commission (FCC). Retrieved August 9, 2023, from https://broadbandmap.fcc.gov
22. FCC. (2022, November 23). *FCC Speed Test App FAQs*. Federal Communications Commission (FCC). https://www.fcc.gov/consumers/guides/fcc-speed-test-app-tip-sheet
23. FCC. (2023, January 8). *Measuring Broadband America*. Federal Communications Commission (FCC). https://www.fcc.gov/general/measuring-broadband-america
24. Feamster, N., & Livingood, J. (2020). Measuring internet speed: current challenges and future recommendations. *Communications of the ACM*, 63(12), 72-80.
25. Fixler, D., Gindelsky, M., & Johnson, D. (2020). *Measuring Inequality in the National Accounts (Updated December 2020)* (Working Paper WP2020-3; BEA Working Paper Series). Bureau of Economic Analysis.

https://www.bea.gov/system/files/papers/measuring-inequality-in-the-national-accounts_0.pdf
26. Frias, Z., Mendo, L., & Oughton, E. J. (2020). How Does Spectrum Affect Mobile Network Deployments? Empirical Analysis Using Crowdsourced Big Data. *IEEE Access*, *8*, 190812–190821. https://doi.org/10.1109/ACCESS.2020.3031963
27. Graham, D. A. (2014, March 27). Rumsfeld's Knowns and Unknowns: The Intellectual History of a Quip. *The Atlantic*. https://www.theatlantic.com/politics/archive/2014/03/rumsfelds-knowns-and-unknowns-the-intellectual-history-of-a-quip/359719/
28. ITU-R. (2015). *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond* (Recommendation ITU-R M.2083-0; M Series). International Telecommunications Union - Radiocommunication Sector (ITU-R). https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf
29. ITU-T. (2001). *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs* (Recommendation P.862 (02/01)). International Telecommunications Union - Telecommunication Standardization Sector (ITU-T). https://www.itu.int/rec/T-REC-P.862-200102-I/en
30. ITU-T. (2019). *Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks* (Recommendation E.806 (06/19)). International Telecommunications Union - Telecommunication Standardization Sector (ITU-T). https://www.itu.int/rec/T-REC-E.806-201906-I/en
31. Johnson, A. (2022, February 2). What's the difference between 5G, 5G Plus, 5G UW, and 5G UC? *The Verge*. https://www.theverge.com/22914668/5g-uw-uc-icon-meaning-verizon-att-t-mobile
32. Karkhanechi, H. M., & Soderstrand, M. A. (1997). Voice quality of cellular mobile phones. *Proceedings of 40th Midwest Symposium on Circuits and Systems. Dedicated to the Memory of Professor Mac Van Valkenburg*, *1*, 485–488. https://doi.org/10.1109/MWSCAS.1997.666179
33. Krasniqi, F., Gavrilovska, L., & Maraj, A. (2019). The Analysis of Key Performance Indicators (KPI) in 4G/LTE Networks. In: Poulkov, V. (eds) Future Access Enablers for Ubiquitous and Intelligent Infrastructures. FABULOUS 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 283. Springer, Cham. https://doi.org/10.1007/978-3-030-23976-3_25
34. Lamberti, A. (2022, April 7). How to Measure VoIP Quality & MOS Score (Mean Opinion Score). *Obkio Blog*. https://obkio.com/blog/measuring-voip-quality-with-mos-score-mean-opinion-score/
35. Lehr, W. (2022). The Changing Context for Broadband Evaluation. In K. Mossberger, E. W. Welch, & Y. Wu (Eds.), *Transforming Everything?* (pp. 25–44). Oxford University Press. https://doi.org/10.1093/oso/9780190082871.003.0002
36. Lehr, W., Heikkinen, M., Clark, D. D., & Bauer, S. (2011). *Assessing Broadband Reliability: Measurement and Policy Challenges*. TPRC39: Telecommunications Policy Research Conference, Alexandria, VA. http://ssrn.com/abstract=1979746

37. Lehr, W., Queder, F., & Haucap, J. (2021). 5G: A new future for Mobile Network Operators, or not? *Telecommunications Policy*, *45*(3), 102086. https://doi.org/10.1016/j.telpol.2020.102086
38. Lehr, W., Smith-Grieco, T., & Woo, G. R. (2008). *Broadband Metrics Best Practices: Review and Assessment*. Massachusetts Institute of Technology. http://people.csail.mit.edu/wlehr/Lehr-Papers_files/Broadband%20Metrics%20Best%20Practices%20Survey%20Feb08.pdf
39. Lehr, W. & Stocker, V. (2023). Next-generation networks: Necessity of edge sharing. Front. Comput. Sci. 5:1099582. https://doi.org/10.3389/fcomp.2023.1099582
40. Macdonald, K. (2023, February 16). European Commission Uses Ookla Data to Demonstrate Territorial Disparities and the Digital Divide in the EU. *Ookla Insights Articles*. https://www.ookla.com/articles/european-commission-uses-ookla-data-eu-digital-divide
41. MÁSMÓVIL. (2023). *Información de Calidad de Servicio de MÁSMÓVIL*. https://www.masmovil.es/static/pdf/calidad-servicio-mm.pdf
42. MINECO. (n.d.). *Documentación de referencia*. Ministerio de Asuntos Económicos y Transformación Digital (MINECO). Retrieved August 9, 2023, from https://avancedigital.mineco.gob.es/es-ES/Servicios/CalidadServicio/DocRef/Paginas/Documentaciondereferencia.aspx
43. Möller, S., & Raake, A. (Eds.). (2014). *Quality of Experience: Advanced Concepts, Applications and Methods*. Springer International Publishing. https://doi.org/10.1007/978-3-319-02681-7
44. Mouly, M., & Pautet, M.-B. (1992). *The GSM System for Mobile Communications*. Bay Foreign Language Books.
45. Movistar. (2023). *Información de Calidad de Servicio de Movistar*. https://www.telefonica.es/es/wp-content/uploads/sites/10/2021/11/Calidad_de_Servicio_Movistar.pdf
46. Mozilla. (n.d.). *Mozilla Location Services*. Retrieved August 10, 2023, from https://location.services.mozilla.com/
47. Ofcom. (2015). *Voice Coverage Obligation Notice of Compliance Methodology* [Statement]. Office of Communications (Ofcom). https://www.ofcom.org.uk/__data/assets/pdf_file/0027/82377/voicecov-compliance-final-20150130.pdf
48. Ookla. (n.d.-a). *About Ookla*. Ookla. Retrieved August 9, 2023, from https://www.ookla.com/about
49. Ookla. (n.d.-b). *Ookla Wind*. Ookla. Retrieved August 9, 2023, from https://www.ookla.com/wind
50. Ookla. (n.d.-c). *Speedtest by Ookla—The Global Broadband Speed Test*. Speedtest.Net. Retrieved August 9, 2023, from https://www.speedtest.net/
51. OpenSignal. (2021, May 26). Understanding mobile network experience: What do Opensignal's metrics mean? *Opensignal*. https://www.opensignal.com/2021/05/26/understanding-mobile-network-experience-what-do-opensignals-metrics-mean
52. OpenSignal. (2023). *Methodology Overview: How Opensignal measures mobile network experience*. https://www.opensignal.com/methodology-overview

53. Orange. (2023). *Informe Trimestral Calidad de Servicio de Orange: Primer trimestre de 2023*. https://www.orange.es/static/pdf/2023T1_InformeCalidadORANGE.pdf
54. Osiptel. (n.d.). *Señal Osiptel* [Map]. Organismo Supervisor de Inversión Privada en Telecomunicaciones (Osiptel). Retrieved August 9, 2023, from https://serviciosweb.osiptel.gob.pe/CoberturaMovil/
55. Principado de Asturias. (n.d.). *Asturias* [Map]. Administración del Principado de Asturias. Retrieved August 9, 2023, from https://sigvisor.asturias.es/SITPA/?webmap=909bf13f608f4af3b776bc5ca0782589&locale=es
56. Principado de Asturias. (2022, October 13). El estudio del Principado sobre conectividad móvil revela que el 92% de la población asturiana tiene buena o excelente cobertura. *Principado de Asturias Actualidad*. https://actualidad.asturias.es/-/el-estudio-del-principado-sobre-conectividad-m%C3%B3vil-revela-que-el-92%25-de-la-poblaci%C3%B3n-asturiana-tiene-buena-o-excelente-cobertura-1?redirect=%2F
57. Proietti, P., Sulis, P., Perpiña, C. C., Lavalle, C., Aurambout, J. P., Batista, E. S. F., Bosco, C., Fioretti, C., Guzzo, F., Jacobs, C., Kompil, M., Kucas, A., Pertoldi, M., Rainoldi, A., Scipioni, M., Siragusa, A., Tintori, G., & Woolford, J. (2022). *New perspectives on territorial disparities* (EUR 31025 EN). Publications Office of the European Union. https://doi.org/10.2760/847996
58. Rohde & Schwarz. (2023). *Measurements that Simulate Human Hearing – Psychoacoustics in ITU-T P.863*. https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/Rohde-Schwarz_Measurements-that-simulate-human-hearing_fly_en_5216_0721_32_v0200.pdf
59. Salter, A. (2023). Cisco announces intent to acquire SamKnows. *SamKnows*. Retrieved August 9, 2023, from https://www.samknows.com/blog/cisco-acquisition
60. Schwartz, B. (2015). The paradox of choice, in J. Stephen (ed.) *Positive psychology in practice: Promoting human flourishing in work, health, education, and everyday life*, Wiley & Sons, New York, 121-138. https://doi.org/10.1002/9781118996874.ch8
61. Shukla, S., & Stocker, V. (2019). *Navigating the Landscape of Programmable Networks: Looking beyond the Regulatory Status Quo*. TPRC47: Telecommunications Policy Research Conference, Washington, DC. https://ssrn.com/abstract=3427455
62. Stocker, V., & Lehr, W.H. (2022). Regulatory Policy for Broadband: A Response to the 'ETNO Report's' Proposal for Intervention in Europe's Internet Ecosystem. https://ssrn.com/abstract=4263096
63. Stocker, V., Lehr, W., & Smaragdakis, G. (2023). COVID-19 and the Internet: Lessons learned. In: Whalley, J., Stocker, V. and Lehr, W.H. (eds.), Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet . Emerald Publishing Limited, 17-69.
64. Stocker, V., Smaragdakis, G., & Lehr, W. (2020). The state of network neutrality regulation. *ACM SIGCOMM Computer Communication Review*, *50*(1), 45–59. https://doi.org/10.1145/3390251.3390258
65. Stocker, V., & Whalley, J. (2018). Speed isn't everything: A multi-criteria analysis of the broadband consumer experience in the UK. *Telecommunications Policy*, *42*(1), 1–14. https://doi.org/10.1016/j.telpol.2017.06.001

66. Stocker, V., & Whalley, J. (2019). Who replies to consultations, and what do they say? The case of broadband universal service in the UK. *Telecommunications Policy*, *43*(8), 101823. https://doi.org/10.1016/j.telpol.2019.04.007
67. SwissQual (2020a), "Facebook Test. Application Note. Release 16.3". Qualipoc user manual.
68. SwissQual (2020b), "NQView, Manual, Application Note. Release 16.1". NQView user manual.
69. Sundaresan, S., De Donato, W., Feamster, N., Teixeira, R., Crawford, S., & Pescapè, A. (2011). Broadband internet performance: a view from the gateway. *ACM SIGCOMM computer communication review*, 41(4), 134-145. doi:10.1145/2043164.2018452
70. Tene, O., & Culnan, M. (2022). *Privacy Metrics Report*. Future of Privacy Forum. https://fpf.org/wp-content/uploads/2022/03/FPF-PrivacyMetricsReport-R9-Digital.pdf
71. The Nielsen Company. (n.d.). *Nielsen*. Nielsen. Retrieved August 9, 2023, from https://www.nielsen.com/
72. Unwired Labs. (n.d.). *OpenCellID*. https://opencellid.org/#zoom=16&lat=37.77889&lon=-122.41942
73. Vodafone. (2023). *Información de Calidad de Servicio de Vodafone*. https://www.vodafone.es/c/statics/informaci%C3%B3n-de-calidad-de-servicio-vodafone-abril-2023.pdf
74. zafaco. (n.d.). *Breitbandmessung Jahresberichte*. breitbandmessung.de. Retrieved August 9, 2023, from https://breitbandmessung.de/ergebnisse

Appendix

Table 1: Comparison of MCS Methodologies of Ookla and OpenSignal

	OOKLA	OPENSIGNAL
DATA COLLECTION		
Type of Measurement	• Active measurements for speed (user-initiated, in foreground) • Passive measurements for coverage (background)	• Active speed test and video test (fixed time period to avoid the ramp up period.). • Passive background measurements (majority of the data collected)
Servers	• Own server network • Multiserver testing	• Content Delivery Networks of popular services, such as Google, Akamai or Amazon
DATA PROCESSING		
Post processing	• Remove tests with potential bias (e.g., excluding including tests performed in controlled environments by network engineers). • Averaging results to create a single sample that summarizes user's internet experience for a time period and geographic area.	• Filtering special events (change of networks, simultaneous call, etc.) and outliers • Averaging results per device: "one device, one vote"
ANALYTICS		
Coverage	**Availability (%)** % users who spent the majority of their time connected to any cellular technology. **Coverage Score™ (1-1000),** based on • **Footprint:** % locations with coverage where the operator has coverage) • **Service:** prob. of access to 4G services	**Availability (%)** % time users had access to a 3G/4G/5G network **Reach (0-10)** Proportion of locations where users were connected to 3G/4G/5G network relative to all the locations visited by the user (in a time frame). **Coverage experience (0-10)** Proportion of (MNO's) locations with 4G/5G coverage relative to locations covered by any MNO
Speed	(Measured) **Speed** (Mbps) • Median Download • Median Upload **Speed Score ™** Combination of 10th, 50th and 90th percentiles in a weight average using 1:8:1 ratio.	Speed **experience** (Mbps) • Average Download • Average Upload

Quality	**Consistency** • Mobile: DL>5Mbps; UL>1Mbps. • 5G and fixed: DL>25 Mbps, UL> 3 Mbps	**(Core / Excellent) Consistent Quality** (% locations that comply), based on 6 KPIs • Average DL speed (> 5 Mbps / 1.5 Mbps) • Average UL speed (>1.5 Mbps ; >500 Kbps) • Latency (< 50 ms / <100 ms) • Jitter (<12 ms / <20 ms) • Packet los (<1%) • Time to first byte
Application Experience	**Applications experience** • Video testing using Speedtest.	**Applications experience (0 – 100 pts)** • Games experience • Mobile video experience • Live video experience • Voice app calling experience • Group video calling experience Using KPIs such as Real-time (videos) • Initial delay • Number of stalling events • Total stalling time • Time spent on each video resolution • Average bitrate at each resolution • Down switching and negative quality changes Non Real-time (Mobile video experience) • Picture quality • Video loading time • Buffering • Playback • Stalling occurrence

Source: Authors based on Connelly (2021) and OpenSignal (2023).