serco
INSTITUTE

User Experience in Government Services:

The Need for a Unique Approach

2022

experienceLab
powered by serco

serco
INSTITUTE

The Serco Institute is an international think tank working to help governments develop the next generation of public service solutions for citizens.

Throughout the Middle East we work collaboratively with government, the private sector, academia, and with citizens to develop the future of public service policy, design and implementation.

info@sercoinstitute.com

experienceLab
powered by serco

Serco launched ExperienceLab in the Middle East in 2020, and as a public services company that manages people, assets and data, ExperienceLab acts as a wraparound for all of our services, focusing on service excellence and the end-user experience.

ExperienceLab helps clients to use the unique understanding of data and insights to design, evolve and adapt services to cater to the end user and deliver world class experiences.

info@serco.ae

We hope this research will guide governments and policymakers as they seek to improve the experience of public service users and gather data on levels and drivers of user satisfaction across the many services provided to citizens and residents.

Outline –
Measuring user satisfaction
with public services

In this report, the Serco Institute and ExperienceLab will examine the issue of measuring user satisfaction – commonly also known as customer or client satisfaction – with public services.

We identify and examine the particular challenges faced by governments in measuring user satisfaction, explore why such efforts are important to public service delivery, provide an overview of existing practices around the world and make the case for a unique approach to measuring user satisfaction with government services rather than using measures designed for other fields.

We hope this research will guide governments and policymakers as they seek to improve the experience of public service users and gather data on levels and drivers of user satisfaction across the many services provided to citizens and residents.

This paper is the first instalment in a series of reports examining how to best measure user experience in the government services space. It outlines the key questions and considerations. Future reports will outline research we will undertake into what tests in the real world tell us about deploying a new, bespoke satisfaction measure to government services.

Methodology

We conducted a thorough literature review of user satisfaction in relation to government/public services around the world.

This review, seeking to outline current thought on best practice in user satisfaction measurement, encompassed academic works by public administration scholars, summaries of well-respected customer satisfaction indices, thought pieces by private sector companies and user experience practitioners, and government website pieces outlining ambitions for greater customer satisfaction.

We identified recurring themes and challenges unique to measuring user satisfaction with government services and several global case studies which we felt illustrated the state of user satisfaction measurement and ambitions in those countries.

Our analysis is also informed by insights from user experience experts who are actively involved in developing and putting into practice new and innovative ways of measuring citizen and resident satisfaction with government services.

Contents

Most existing measures of customer satisfaction are developed and intended for use in private sector contexts and focused on sales or customer service.

▮▮ | | ▮ Executive summary

Governments around the world have taken an increasing interest in measuring the satisfaction of citizens and residents with the services they provide. Given it is a government's duty to provide its people with the services and the means to go about their lives, this is perhaps unsurprising.

However, most existing measures of customer satisfaction are developed and intended for use in private sector contexts and focused on sales or customer service, such as retail or ecommerce. Deploying commercially-focussed models of measuring satisfaction can be problematic in the context of government services, because: these services often operate in a far less competitive environment; governments have an obligation to provide services for all citizens and residents and so must account for a much wider range of demographic, language and accessibility considerations; and governments often also provide services which are unpopular regardless of whether they are efficient, well-managed or deliver a good user experience (such as tax collection).

Through an in-depth exploration of existing literature alongside insights from user experience experts, this paper offers an in-depth exploration of the challenge of measuring user satisfaction with government services.

We look at how measuring user satisfaction with government services is impacted by a range of different external factors, and the need to distil findings into something which may easily be acted upon by policymakers. We also explore case studies from around the world, examining how governments have sought to overcome the challenge of understanding the feelings of their citizens and residents towards services. We also attempt to distil the key considerations of how to build an effective government services user satisfaction measure into two key parts: Data collection (or **inputs**) and data analysis (or **outputs**).

Inputs

Are the 'things' (such as the actions, words, and characteristics) which must be captured in order to gauge user satisfaction with services. We re-examine long-held beliefs about current user satisfaction models by asking: How can we complement easier-to-digest quantitative measures, such as rating services on a number scale, with potentially more insightful qualitative data uncovering what commonly drives a user's satisfaction? Should we be asking service users to state their preferences, or should we be tracking their behaviour to reveal their true preferences? And at which point in time during a user's interaction with a government service should we be measuring their satisfaction?

Outputs

Are how we translate the insights uncovered by measuring inputs into actionable recommendations for policymakers. We again seek to probe existing beliefs about best practice by asking: What balance do we strike between insight-rich qualitative findings and more easily discernible quantitative data? How can we create a comparable series of metrics of public services to measure changes in satisfaction over time? And should we rely more on descriptive statistics, focused on identifying the visible trends and characteristics exhibited by a sample, or on inferential statistics, drawing predictions based on a sample about a broader population which may yield more valuable insights into how the general population feel about a public service?

We conclude by arguing that to properly measure people's sentiments towards the government services they use, a new approach is needed, one which takes into account the uniquely broad public service user base and the inherent challenges of measuring something as intangible as satisfaction. This paper identifies many of the problems with existing methods of measuring satisfaction and forms a fundamental building block as the basis for a new government services-oriented metric of user satisfaction.

Why governments must measure user satisfaction

1. Why do governments need to measure user satisfaction?

2. Are some services pre-disposed to better satisfaction scores?

3. What is the unique challenge of accurately measuring satisfaction with public services?

Summary

As in the private sector, governments are increasingly interested in measuring how satisfied the users of their services are

However, governments have often copied the measures used by the private sector and tried to fit them to measure satisfaction with public services

Existing research has already found that citizens and residents will always score some public services more highly than others – making it difficult to accurately compare different public services

Furthermore, government services are often accessed for very different reasons to those in the private sector. Private sector service users often have access to multiple options and use only those which they wish to, whereas government services are often used by people who are required to do so – to manage visas or identity documents, for example.

Equally, public services are often delivered through very different structures to those found in the private sector – for example, through a single, monopolistic provider – making the existing ways to measure customer satisfaction even more unsuitable.

Fundamentally, the existing private sector-developed measures of user satisfaction offer a reasonable standard for monitoring how people feel about public services, but they fail to capture many of the unique drivers of satisfaction when it comes to government services.

Why governments must measure user satisfaction – the past & present

In the past few decades there has been a growing consensus that public services must be centred not around government needs, but those of the citizens and residents who use the services.

As such, the concept of user satisfaction has become of increasing interest to governments, and a number of metrics have been used to measure the performance of public services and the resultant satisfaction among public service users.

The need to measure user experience, or 'customer experience' as it is often known, was primarily driven by private sector organisations in an attempt to encourage sales, boost customer retention and increase brand value. Pioneers of the concept, such as the retailer Harry Gordon Selfridge – who is often credited with coning (or popularising) the phrase 'the customer is always right' – would recognise many of the techniques used by businesses today to try and measure the satisfaction levels of their customers. What is more, a range of techniques first used in the private sector, from satisfaction surveys to mystery shoppers, have become increasingly commonplace in the government services sector.

Why do governments need to measure user satisfaction?

There are a number of reasons why user satisfaction is a matter of interest to governments. The most obvious, and the most simple, is that it is the job of governments to keep it population safe and happy, to provide them with access to the services – such as healthcare, transport, and housing – they need to go about their lives. Public service delivery and performance should, therefore, be centred around and gauged by the citizens and residents who avail themselves of these services. Government services scoring highly on user satisfaction are additionally an important contributing factor to positive perception of the country which provides these services, which in turn results in more tangible benefits: such countries may find it easier to attract foreign talent and investment, as well as tourists. Furthermore, numerous studies have found strong evidence linking service quality and operational performance to customer satisfaction and loyalty [1]. Recognition of this has led, particularly since the 1990s, to a growing focus on customer satisfaction by public service providers.

Given the obligation by governments to serve their population, it is important that satisfaction by government service users is measured, alongside the factors which may influence satisfaction ratings by customers: for instance, historically high levels of tax, especially in the West, has set high expectations from service users who want to see this revenue put to good use.

Understanding such factors is vital to optimising public service delivery. Governments' attempts to measure user satisfaction must also be viewed in the context of greater competition not only with the private sector, but also with other governments, as comparisons between governments' performances are made possible by people's international mobility, experience of other governments and consumption of international news. This is especially true of countries with large expatriate or transient populations, such as the UAE, which has an exceptionally large population of foreign nationals (88%) and one of the world's most advanced user satisfaction, or happiness, programmes [2].

Research furthermore suggests that satisfaction with public services is associated with trust in government: Belgian researchers, for instance, have found that higher satisfaction with public services in their country's Flemish Region is associated with higher trust in Flemish, Belgian and European political institutions [3], while Professor Claes Fornell, publisher of the American Customer Satisfaction Index (ACSI), has suggested a causal relationship between rising ACSI scores in US federal government services and increasing trust in government, as measured by public confidence that 'government agencies will do a good job in the future' [4]. There is, therefore, something of an inherent interest for government in ensuring public service users are satisfied, and measuring user satisfaction should therefore be high-up any government's agenda.

[1] 'The Behavioral Consequences of Service Quality' – Zeithaml et al., Journal of Marketing 60(2); 'The Impact of Operations Performance on Customer Loyalty' – Kumar et al., Service Science 3(2).

[2] 'How the UAE government modernized citizen services' – McKinsey; 'Transformation, digitisation & happiness: Public opinion on UAE government services' – Serco Institute.

[3] 'On the relative role of the public administration, the public services and the political institutions in building trust in government in Flanders' – Jarl K. Kampen et al., Public Performance & Management Review 29(4).; 'Transformation, digitisation & happiness: Public opinion on UAE government services' and 'Transformation, digitisation & happiness: Public opinion on government services in Saudi Arabia' – Serco Institute.

[4] 'ACSI Commentary: Federal Government Scores' – Claes Fornell.
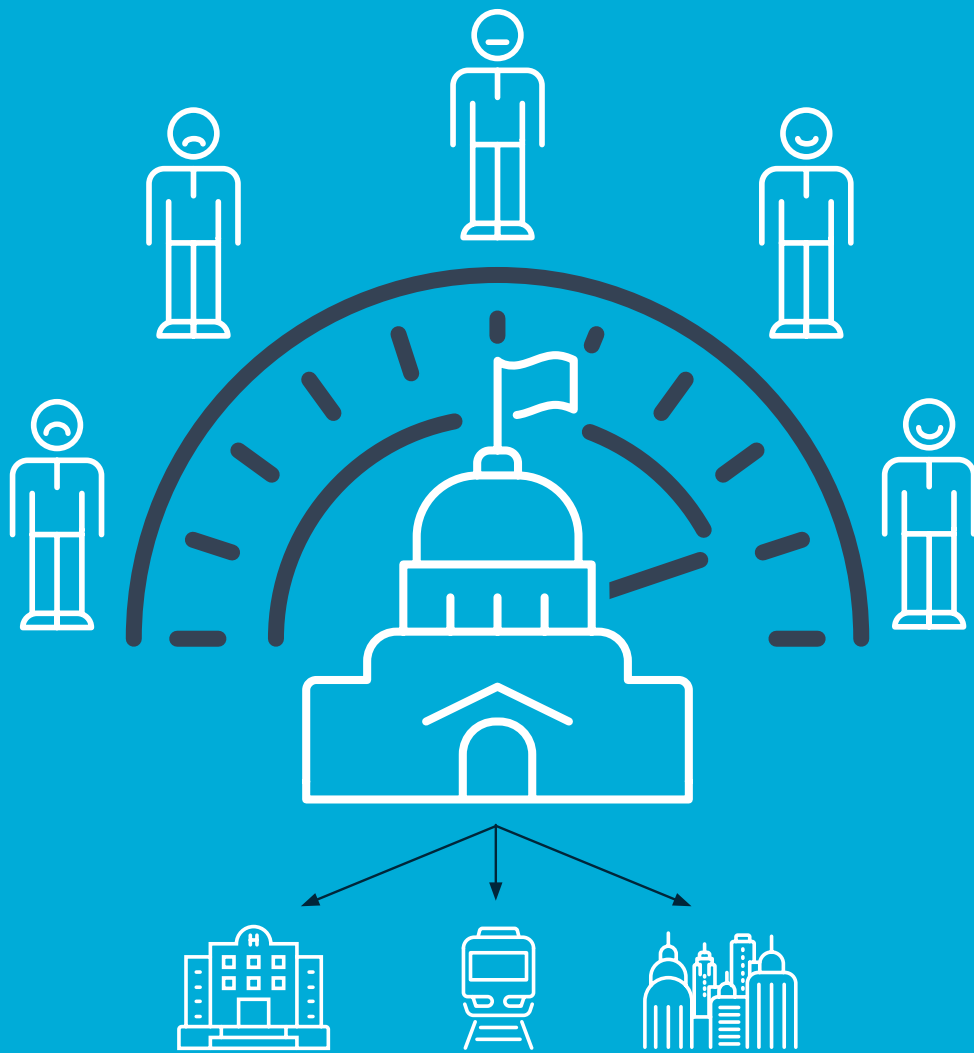
There are a number of reasons why user satisfaction is a matter of interest to governments. The most obvious, and the most simple, is that it is the job of governments to keep it population safe and happy, to provide them with access to the services – such as healthcare, transport, and housing – they need to go about their lives
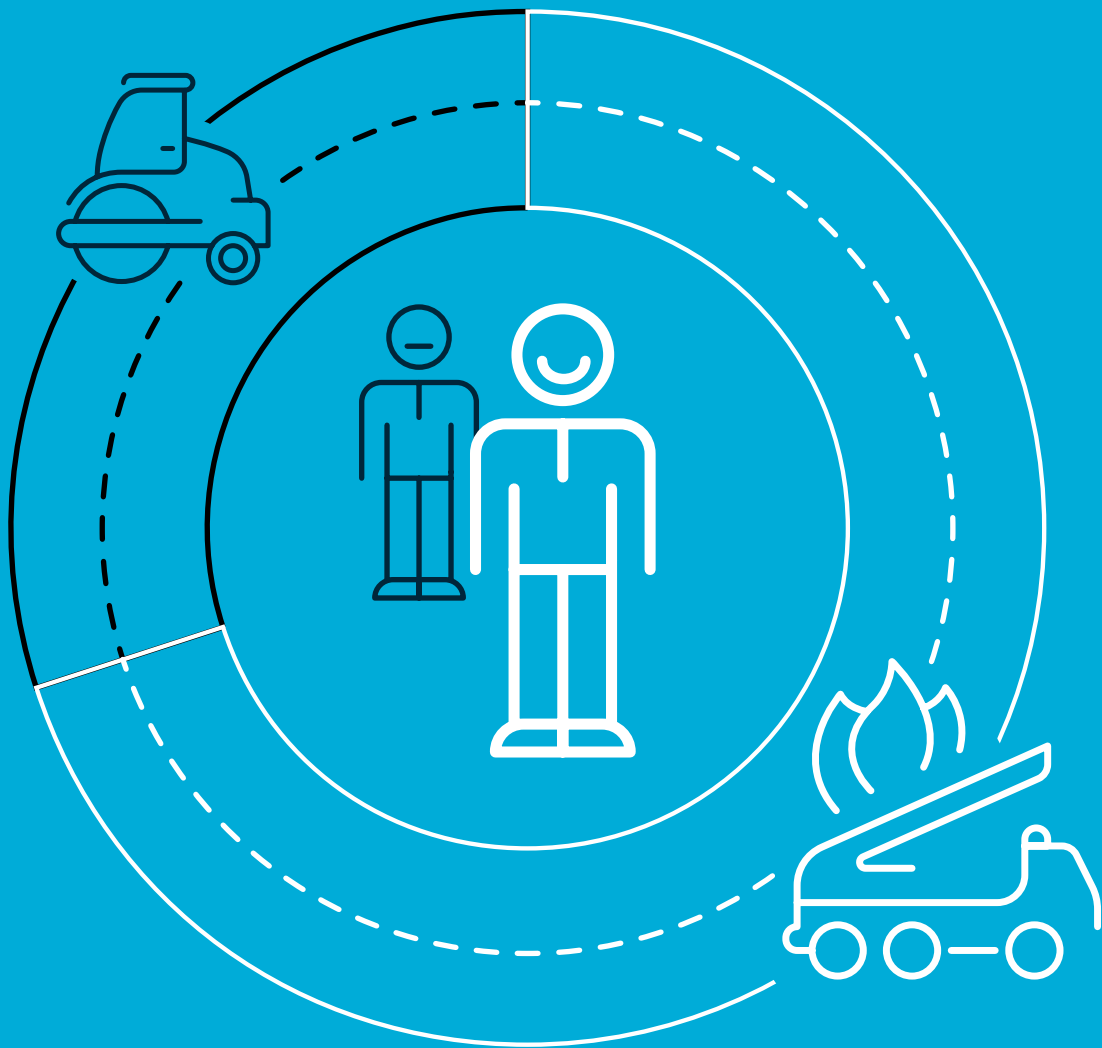
A comparative study assessing satisfaction with US local government services indicated that firefighting services consistently earned higher satisfaction ratings than road repair services [5].

Are some services predisposed to better satisfaction scores?

Satisfaction scores should be regarded, however, with a degree of caution, as existing satisfaction metrics largely rely on the use of surveys, which are a subjective means of data collection. Pre-existing attitudes towards government or specific public services may therefore have some bearing on satisfaction judgments, and there is evidence to support this: a comparative study assessing satisfaction with US local government services indicated that firefighting services consistently earned higher satisfaction ratings than road repair services [5]. Additionally, while all public services are inextricably linked to public policy, certain services, such as immigration and asylum, are unlikely to have been experienced by a majority of the population. Satisfaction with these services is therefore considerably more likely to be coloured by an individual's political positions than by experience of using the service. These intrinsic biases are an unquantifiable factor which weakens the analytical value of the survey data with which public service user satisfaction is usually measured.

Furthermore, research suggests that public service users' prior expectations have a significant influence on stated satisfaction levels, a phenomenon referred to as expectancy disconfirmation theory. This effect has been observed in studies of customer satisfaction: A 2006 study into satisfaction with local government services in the US strongly supports expectancy disconfirmation, indicating that the performance of city services had a strong indirect influence on satisfaction by exceeding respondents' prior expectations of the quality of service [6].

Previous experience of a public service is a major factor impacting an individual's expectations, but even here all things are not equal. Numerous studies find that negative experiences of a public service exert a far stronger influence on satisfaction than positive experiences:

A 2004 study measuring drivers of user satisfaction across a range of public services in New York City indicated that while three best-rated public services contributed little to overall perceptions of quality and satisfaction with public services in New York, the three worst-rated services were found to be important drivers of overall quality and satisfaction [7].

Another piece of research, based on data from a 2002 citizen satisfaction survey by the Flemish government, indicates that negative experiences of a public agency have a far bigger impact on trust in public services than positive experiences [8].

All this indicates that measuring satisfaction cannot account for a number of unquantifiable biases and how these biases affect people's perceptions of public services. Satisfaction ratings are not always made based purely on service performance or delivery, and the inherent biases of some respondents should be taken into account when considering survey-based satisfaction scores.

[5] 'Explaining Citizen Satisfaction and Dissatisfaction with Public Services' – Steven van de Walle, in The Palgrave Handbook of Public Administration and Management in Europe.

[6] 'Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government' – Gregg G. Van Ryzin, Journal of Public Administration Research and Theory 16(4).

[7] 'Drivers and Consequences of Citizen Satisfaction: An Application of the American Customer Satisfaction Model to New York City' – Gregg G. Van Ryzin et al., Public Administration Review 64(3).

[8] 'Assessing the relationship between satisfaction with public service delivery and trust in government' – Jarl K. Kampen et al., Public Performance & Management Review 29(4).

What is the unique challenge of accurately measuring satisfaction with public services?

Most measures for user satisfaction were designed to be deployed in commercial settings. Common metrics like Customer Satisfaction Scores (CSAT), Net Promoter Scores, or Customer Effort Scores were designed to be used by businesses with the ultimate aim of enhancing their commercial positions.

In particular – perhaps unsurprisingly – sectors heavily focussed on sales or customer service have often led the way in this area. From in-person retail to eCommerce, these services are, by definition, transactional, commercial and more often than not in a competitive market environment. Government services, on the other hand, are not necessarily transactional, rarely commercial, and more frequently delivered through a single, monopolistic structure.

Furthermore, government services are not necessarily accessed as a result of an individual's 'want', but rather as a result of 'need' – be it a legal requirement, a financial necessity or in search of support. This means there is often no 'self-selecting' consumer of public services – like there would be in most commercial settings where people only choose to access a service as they want to obtain something, not because they must. Furthermore, this means almost all services need to be accessible by all citizens and residents. Therefore, the range of different demographics, languages and access needs for public services is larger than any commercial equivalent.

These factors mean that

Measuring user satisfaction with public services is a unique challenge. However, more often than not, the solutions employed are transposed from commercial settings.

Furthermore, the expectations and drivers of satisfaction amongst government service users as compared to customers in a commercial setting are very different. As noted above, on the most fundamental of levels, disaggregating a government service user's sentiment from their feelings towards the policy which is driving the outcome of the service is often impossible. For example, a person whose visa application has been rejected through an online portal is unlikely to give an objective view of how satisfied they are with their experience of using the digital system. Moreover, a service user might harbour negative feelings towards a government service as a result of their political leanings and coloured by the amount of taxes they pay.

Of course, people's expectations of customer experience are also impacted by the improvements championed in the private sector. It is, after all, the same people who use Uber, Deliveroo and Airbnb, who also use government services. The unique nature of public services does not mean that they should be insulated from innovations in improving user experience. What needs to be explored, however, is how we can measure the experience of government service users in a way which accommodates their different characteristics.
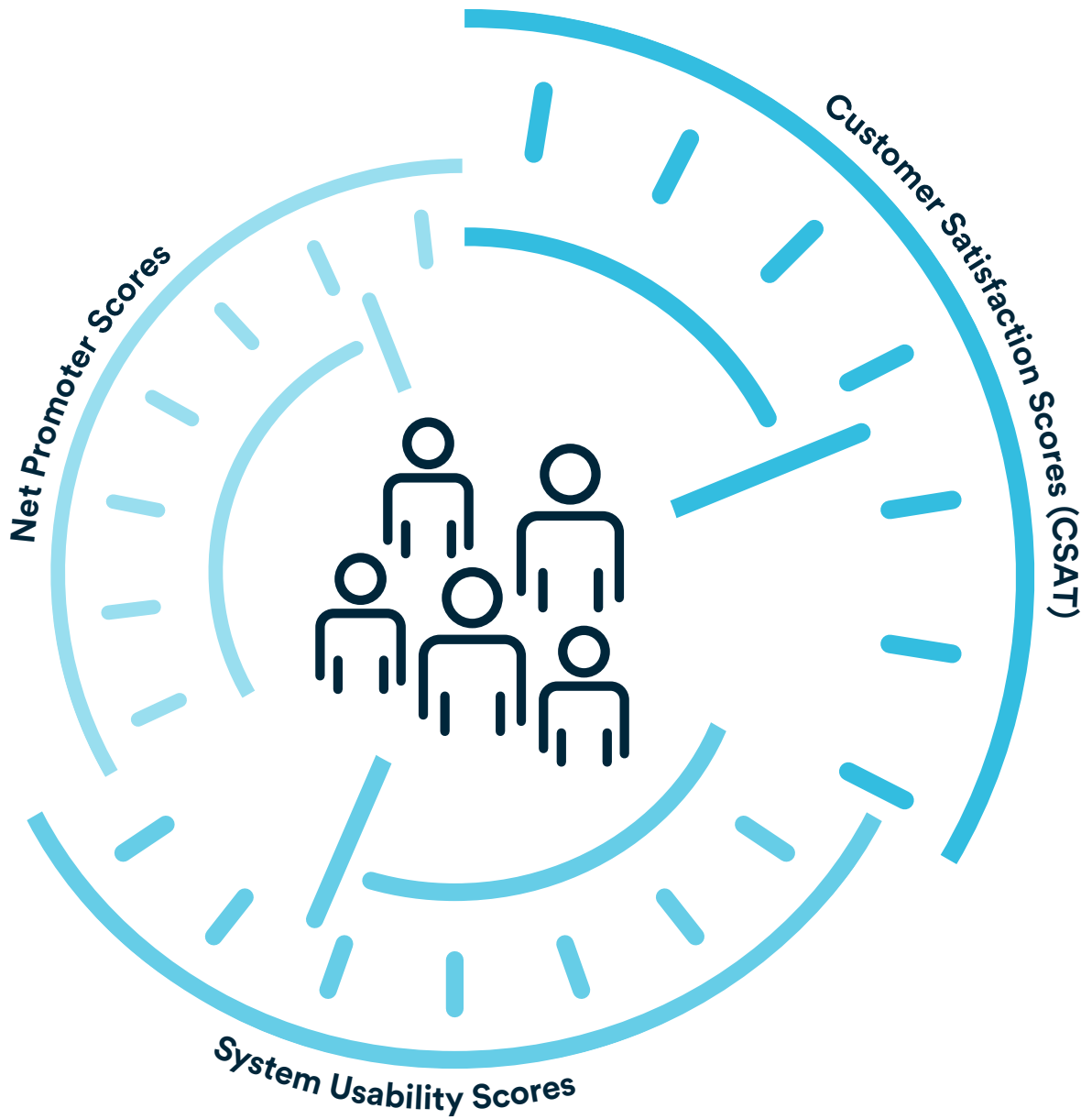
Customer Satisfaction Scores (CSAT)

Net Promoter Scores

System Usability Scores

User satisfaction measurement has become increasingly popular among governments around the world, with several countries, such as the UAE, officially incorporating improvement to public service user experience into their agendas.

Case studies –
How governments are already measuring user satisfaction with services

Countries all over the world have increasingly invested in ways to measure user satisfaction.

North America

Europe

Middle East

Asia-Pacific

Summary

||| |

North America

A huge range of service-specific, state-level and federal initiatives exist to measure government service user satisfaction in the US. However, one of the most well-known and prolific is the American Customer Satisfaction Index (ACSI). As of 2021 the ACSI also measures citizen satisfaction through interviews with public service users, analysing responses in relation to over 100 services, programmes and websites of federal agencies in the United States. A system developed in Canada, the Common Measurement Tool, is now used by many public sector organisations in the country at the municipal, provincial/territorial and federal level to measure user satisfaction and expectations, and identify areas for improvement in public services.

||| |

Europe

User satisfaction with public services in Europe is measured by national governments (and at a sub-state, regional level), as well as the European level. Perhaps most notably, the first index used to measure customer satisfaction with public services, and which went on to form the basis of the American Customer Satisfaction Index, was developed in Sweden – Statistics Sweden continues to conduct yearly citizen surveys.

||| |

Middle East

Saudi Arabia (KSA) has also been investing in measuring government service user experience. In February 2019, the KSA Government launched the Watani smartphone app, allowing KSA citizens, residents and visitors to rate public services and provide performance feedback on 80,000 services offered by around 30,000 government service centres around the Kingdom. The UAE paid particular attention to user satisfaction with government services developing its world-leading National Programme for Happiness and Wellbeing. Programmes include the deployment of 'Happiness Meters' in all federal entities, which allow service users to register how they feel about the service they use in real-time.

||| |

Asia-Pacific

Both Australia and New Zealand have run national user satisfaction surveys to gauge citizens' satisfaction with public service. A national survey in Australia, as well as state-level surveys, see public sentiment measured on a repeated basis every year. On New Zealand, the government uses a modified version of the Canadian, Common Measurement Tool to carry out an annual survey on public services.

Case studies –
How governments are already measuring user satisfaction with services

User satisfaction measurement has become increasingly popular among governments around the world, with several countries, such as the UAE, officially incorporating improvement to public service user experience into their agendas. In Western countries, notable examples of user satisfaction metrics include the American Customer Service Index (ACSI), which as of 2021 measures citizen satisfaction with over 100 services, programmes and websites of federal agencies in the United States [9], and the Common Measurements Tool, developed in Canada but also used widely in New Zealand to measure customer satisfaction with the quality of government services in those countries [10]. Similar satisfaction surveys are also conducted by governments at the subnational level: for example, the state government of South Australia runs an annual customer satisfaction survey to measure the performance of their public services, in comparison to other jurisdictions such as New South Wales, Victoria and Queensland [11].

[9]   'ACSI Federal Government Report 2020' –
      American Customer Satisfaction Index.

[10] 'A guide to using the Common Measurements
      Tool' – New Zealand Government.

[11] 'Government Customer Satisfaction Survey' –
      Government of South Australia.

Middle East

Saudi Arabia has demonstrated a similar ambition to develop tools to measure satisfaction among users of public services. The National Center for Performance Measurement, or Adaa, is a government body aiming to establish unified tools to measure performance in Saudi public entities and drive up performance and efficiency [12].

In February 2019, it launched the Watani smartphone app, allowing KSA citizens, residents and visitors to rate public services and provide performance feedback on 80,000 services offered by around 30,000 government service centres around the Kingdom [13]. Improving and developing an index to measure user satisfaction across public services is also mentioned in the KSA Government's strategic development plan, Vision 2030 [14].

Elsewhere in the Middle East, the United Arab Emirates (UAE) has become a world leader in measuring customer/user satisfaction, having made 'happiness' an area of high priority in 2016. Through its National Programme for Happiness and Wellbeing, the UAE Government has pledged to incorporate happiness into all functions, policies and services of government, overseen by a Minister of State for Happiness [15].

Critically, this Programme includes a plank for the development of tools to measure UAE residents' happiness: this aspiration was achieved by the launch of the Happiness Meter, a feedback system for UAE residents who have just used a public service to indicate their satisfaction on a sliding scale. By April 2019, all federal government entities in the UAE had deployed the Happiness Meter. Customer service centres, known as 'Customer Happiness Centers', are equipped with iPads for customers to complete a Happiness Survey. The Meter is also integrated into government websites and apps, including the Federal Government services portal, Khadamati, allowing UAE residents to report on their experiences of the 2,500 services offered by the Federal Government. The Happiness Meter appears popular: in an April 2019 survey, 90% of respondents considered it a success, and in 2018 it recorded a 92.6% happiness rating [16].

The UAE Government is working to expand options for public service users to return feedback: in March 2020, it launched 'UAE Mystery Shopper', an app allowing UAE residents to rate government services and submit feedback on government offices [17].

[12] 'About Adaa'
– The National Center for
Performance Measurement.

[13] 'Smartphone application launched to improve
public agencies' services in Saudi Arabia' –
Arab News.

[14] 'Assessing Saudi Vision 2030: A 2020 Review'
– Atlantic Council.

[15] 'Happiness' – UAE Government.

[16] 'GX Case Study: The UAE's Happiness Meter'
– Government Experience.

[17] 'UAE Cabinet launches 'Mystery Shopper'
app to rate government services'
– The National.

North America

One of the most prominent metrics for measuring user satisfaction with public services is the American Customer Satisfaction Index (ACSI). ACSI scores are calculated as the weighted average of three factors: customer expectations of the quality of the service being provided; perceived quality, or customers' assessment of the quality or performance of the service based on recent consumption experience; and perceived quality, or the perceived quality or performance of services based on price paid [18].

Developed originally for and still used by companies, as of 2021 the ACSI also measures citizen satisfaction through interviews with public service users, analysing responses in relation to over 100 services, programmes and websites of federal agencies in the United States. It produces an annual Federal Government Report and has been selected by the US Federal Government as a standard metric for measuring user satisfaction [19].

The Common Measurements Tool (CMT) is another survey-based metric designed to gauge satisfaction with public services, which identifies elements influencing client service – user expectations, perceptions, satisfaction levels, importance and priorities for improvement – and asks public service users to rate their experiences across five dimensions of service delivery: responsiveness; reliability; access & facilities; communications; and cost [20].

Developed in Canada, the CMT is now used by many Canadian public sector organisations at the municipal, provincial/territorial and federal level to measure user satisfaction and expectations, and identify areas for improvement in public services [21].

It is also in use in New Zealand, where the Public Service Commission, formerly the State Services Commission, has adapted the CMT for use in the quarterly Kiwis Count survey, which measures New Zealanders' trust in and satisfaction with public services in their country [22].

[18] 'The Science of Customer Satisfaction' – American Customer Satisfaction Index.

[19] 'ACSI Federal Government Report 2020' – American Customer Satisfaction Index.

[20] Client Satisfaction Surveying: Common Measurements Tool' – Government of Canada Publications.

[21] About the Common Measurements Tool' – CitizenFirst.

[22] 'A guide to using the Common Measurements Tool' – State Services Commission.

Asia-Pacific

Both Australia and New Zealand have run national user satisfaction surveys to gauge citizens' satisfaction with public services: as mentioned above, the survey in New Zealand is known as Kiwis Count and takes place quarterly. Since March 2019, a similar survey has been run by the Australian Government, reviewing Australians' trust in and satisfaction with the Australian Public Service (APS) and tracking how these sentiments change over a series of quarterly 'waves'. Surveyed aspects of service delivery include ease of use, staff friendliness, treatment and time spent in reaching an outcome [23].

Several Australian states additionally run annual Customer Satisfaction Measurement Surveys, examining public satisfaction with and perception of public services. These cover public services in more than one jurisdiction: the survey of the Government of South Australia, for instance, surveys users of public services based not only in South Australia, but also in the states of New South Wales, Victoria and Queensland, to compare South Australia's public services against those provided by other Australian states. The survey examines perceptions of government services by comparing prior expectations to service performance, while also measuring factors such as reported satisfaction and preferred means of accessing services.

Notably, it includes and tracks the sentiments not only of individual consumers, but of businesses in these states, and also gathers data on consumers' ideal standard of service against which to benchmark existing services.

By running this survey, the Government of South Australia hopes to monitor changes in expectations and satisfaction levels over time, address user-identified areas for improvement and assess service performance against the criteria of employees, values, process and goals [24].

New South Wales runs a very similar and identically named Customer Satisfaction Measurement Survey, tracking the performance of 22 public services against users' satisfaction, expectations, trust in government, effort in engaging with government services and comparison to users' ideal standard of service. Once again, the New South Wales survey also polls both individual consumers and businesses, and also public service users outside New South Wales in order to compare NSW public services against those in other jurisdictions: generally, surveys additionally poll customers in Victoria, Queensland and South Australia [25], while the 2019 edition also included service users in the UK and New Zealand [26]. Customer satisfaction is said to be of central importance to the New South Wales Government, which has a Department of Customer Service, overseen by a Minister for Customer Service, responsible for the improvement and optimisation of public service and service delivery in the state [27].

[23] 'Citizen Experience Survey'
– Australian Government.

[24] 'South Australian Customer Satisfaction Measurement Survey 2020'
– Government of South Australia.

[25] 'UAE Cabinet launches 'Mystery Shopper' app to rate government services'
– The National.

[26] 'About Adaa'
– The National Center for Performance Measurement.

[27] 'Smartphone application launched to improve public agencies' services in Saudi Arabia' –
Arab News.

Europe

User satisfaction with public services in Europe is measured by national governments, as well as the European level. At a national level, one of the best examples of a satisfaction measurement metric across government services generally is that of the Norwegian Government's Agency for Public and Financial Management, which carries out regular 'population surveys' measuring Norwegians' satisfaction with public services [28]. User satisfaction metrics exist in other European countries, though often on a smaller scale: UK Government websites have functions which allow users to report feedback on their experience, with the Government Digital Service issuing advice to GOV.UK sites regarding the points of a user's journey at which satisfaction should be measured, how satisfaction data should be used and how to continually monitor changes in satisfaction [29]. Most notably, the first index used to measure customer satisfaction with public services, and which went on to form the basis of the American Customer Satisfaction Index, was developed in Sweden [30]– Statistics Sweden continues to conduct yearly citizen surveys [31]. Additionally, some user satisfaction metrics are in use in subnational jurisdictions: in Flanders, Belgium's Dutch-speaking region, ministries and public agencies of the region's government have employed citizen satisfaction surveys to monitor public service delivery [32], as part of the region's 'Flanders-in-Action' initiative to improve the performance of Flemish public sector organisations [33].

At the European level, user satisfaction with public services is usually measured in the form of public opinion on issues and topics in European politics and public policy.

Since 1997, however, Eurobarometer has been measuring opinion on public services, often referred to as services of general interest, in addition to polling on broader public sentiment towards the European Union and EU policy. One recent report, released in September 2021, contains data from all EU member states on respondents' priorities among a field of public service categories, including health, immigration, education and housing, among others, and benchmarking trust in EU institutions against those of the national government [34]. Eurobarometer also conducts surveys on more specific public policy areas, such as the EU's Covid-19 vaccination campaign, as well as surveys asking Europeans to identify those policy areas they believe should be priorities for the EU. Similarly, the European Social Survey examines Europeans' attitudes towards various topics relevant to Europe, such as immigration, health inequalities, justice and climate change. The Survey's findings have been used to gauge satisfaction with public services in different regions and countries, such as by NordMod 2030 which sought to measure satisfaction with public services in the Nordic countries [35].

These case studies illuminate common practices used by governments around the world as they measure user satisfaction with public services. However, for the reasons stated above and as explored further in the following section, we are sceptical as to these practices' ability to fully capture the nature of public sentiment towards government services. A new approach, tailored specifically to measuring satisfaction with public services, is needed.

[28] 'Innbyggerundersøkelsen 2019' – DFØ.

[29] 'Measuring user satisfaction' – GOV.UK.

[30] 'A National Customer Satisfaction Barometer: The Swedish Experience' – Claes Fornell, Journal of Marketing 56(1).

[31] 'SCB:s medborgarundersökning' – SCB.

[32] 'Assessing the relation between satisfaction with public service delivery and trust in government' – Jarl K. Kampen et al., Public Performance & Management Review 29(4).

[33] 'The Call for Open and Innovative Government: An Overview of Country Initiatives' – OECD.

[34] 'Standard Eurobarometer 95 - Spring 2021' – Eurobarometer.

[35] 'Exploring public attitudes, informing public policy' – European Social Survey.

||||

Online access is revolutionising where, how and when people interact with government services.

The need for a new approach – Embracing change in user habits, embracing change in measuring them

1.   Inputs

2.   Outputs

3.   Creating a user satisfaction score fit for public services – Hitting a moving target

Summary

New research, as well as new ways of accessing services is driving new thinking about how we measure user satisfaction.

The ever-increasing number of services being delivered digitally – particularly in the context of the pandemic and post-Covid-19 landscape – is having a profound impact on the ways in which people think about government services and how we can measure how they feel about their experience of using them

Capturing how users really feel about their experience of using a service will require the measurement of a range of different 'inputs'. These inputs could include information from survey questions, but also analysis of other data points, like how long a task takes to complete. Research is key in identifying the inputs which will most accurately measure users' satisfaction with a given service.

We also need to be aware when measuring inputs of the difference between 'stated preference' and 'revealed preference' – sometimes people will say one thing, but their actions will show that they prefer another. That is why a mix of 'active' inputs – where the user gives direct feedback – and 'passive' inputs – where we measure the actions of the user – is important.

Bringing all the inputs together to create an 'output' which policy makers and service designers will make decisions is equally important.

Outputs need to be simple and comparable, but not stripped of so much detail that they lack key insights or could be misinterpreted.

The need for a new approach – Embracing change in user habits, embracing change in measuring them

The way in which users are accessing government services is changing, partly driven by a transformation in how services are delivered and partly as a result of wider shifts in society. Digitisation is one of the most visible and widespread drivers of this change.

Online access is revolutionising where, how and when people interact with government services.

For example, world leader Estonia now claims to have 99% of its public services available online [36]. However, the digitisation of public services must be accompanied by a concerted effort to centre the digital transformation around the needs of users, and have these at the core of service design: a 2020 discussion paper found that the UK Government's digitisation efforts had been limited in their success by a failure to 'genuinely engage citizens' and other stakeholders in the design of digital services [37].

The pandemic has also had a significant impact on how people interact with services, including those delivered by government. Covid-19's long-term effects on the rate of home-working and use of applications such as online meeting services is as yet unknown, but it has already precipitated a greater shift to the use of online services. Popular video-conferencing software Zoom, for example, saw usage grow at an unprecedented rate, from 60 billion annual meeting minutes in the 12 months to February 2019 to a peak of 3.3 trillion in the 12-month period prior to the end of September 2021 [38] [39].

Furthermore, our research has found that Covid-19 has also increased the desire for more public services to be accessible online. A nationally representative Serco Institute survey found that 74% of people in the UAE and 62% in KSA indicated that they were more likely to use digital government services as a result of Covid-19 [40].

Although new measures and tools become available to identify levels of satisfaction as services increasingly move online, the issue of better measurement is not going to be solved by the increased digitisation of services. That is not to say that digitisation cannot help us understand user satisfaction more effectively – as the above case studies illustrate, some of the best practice for measuring user satisfaction is to be found in public services delivered and/or accessed online. The real issue is, however, the embedded approaches to measuring user satisfaction are often not fit for their stated purpose or failing to keep up with the latest research and techniques. Moreover, a significant portion of service users will continue to access services in-person, over the phone or through other non-digital means. In fact, some services simply cannot be delivered digitally.

[36] 'e-Estonia' – Estonian Government.

[37] 'Better Digital Government: Obstacles and Vision' – Commission for Smart Government.

[38] 'Zoom User Stats: How Many People Use Zoom in 2022?' – Backlinko.

[39] 'Transformation, digitisation and happiness: Public opinion on UAE government services' – Serco Institute.

[40] 'Transformation, digitisation & happiness: Public opinion on government services in Saudi Arabia' – Serco Institute.

Therefore, regardless of how sophisticated the technological tools for tracking sentiment, governments will continue to need to employ a range of different methods of data collection and analysis to understand how users feel about the services they are delivering.

The private sector has employed increasingly sophisticated methods to track user sentiments. This is particularly true in the eCommerce and social media sectors. As is often the case, the public sector can learn and borrow from these innovations, but the models designed for the highly transactional systems that are more common in the private sector are not necessarily the best fit for the complex services delivered by governments and their partners. Customer satisfaction measures in the private sector are often focussed around sales and seek to allow comparability with competitors. Government services are, however, not necessarily transactional and are often delivered to citizens and residents in a monopolistic manner – with a single provider and no comparator.

As already noted, measuring satisfaction across public/government services is arguably significantly more complex.

The diversity of its user-base, accessibility requirements and the variety of services governments need to provide all create significant challenges to accurately measuring satisfaction.

Perhaps even more difficult than measuring satisfaction is the distillation of those measurements into something that policymakers – who are rarely user experience experts – can use and base decisions on regarding the design and delivery of services. Many policymakers will have to consider swathes of data, often relating to different services. Highly complex and varied measurements become cumbersome, difficult to understand and in effect unusable, if they are engaged with at all.

Put simply, there are two elements of the 'measurement' conundrum facing those designing and delivering public services:
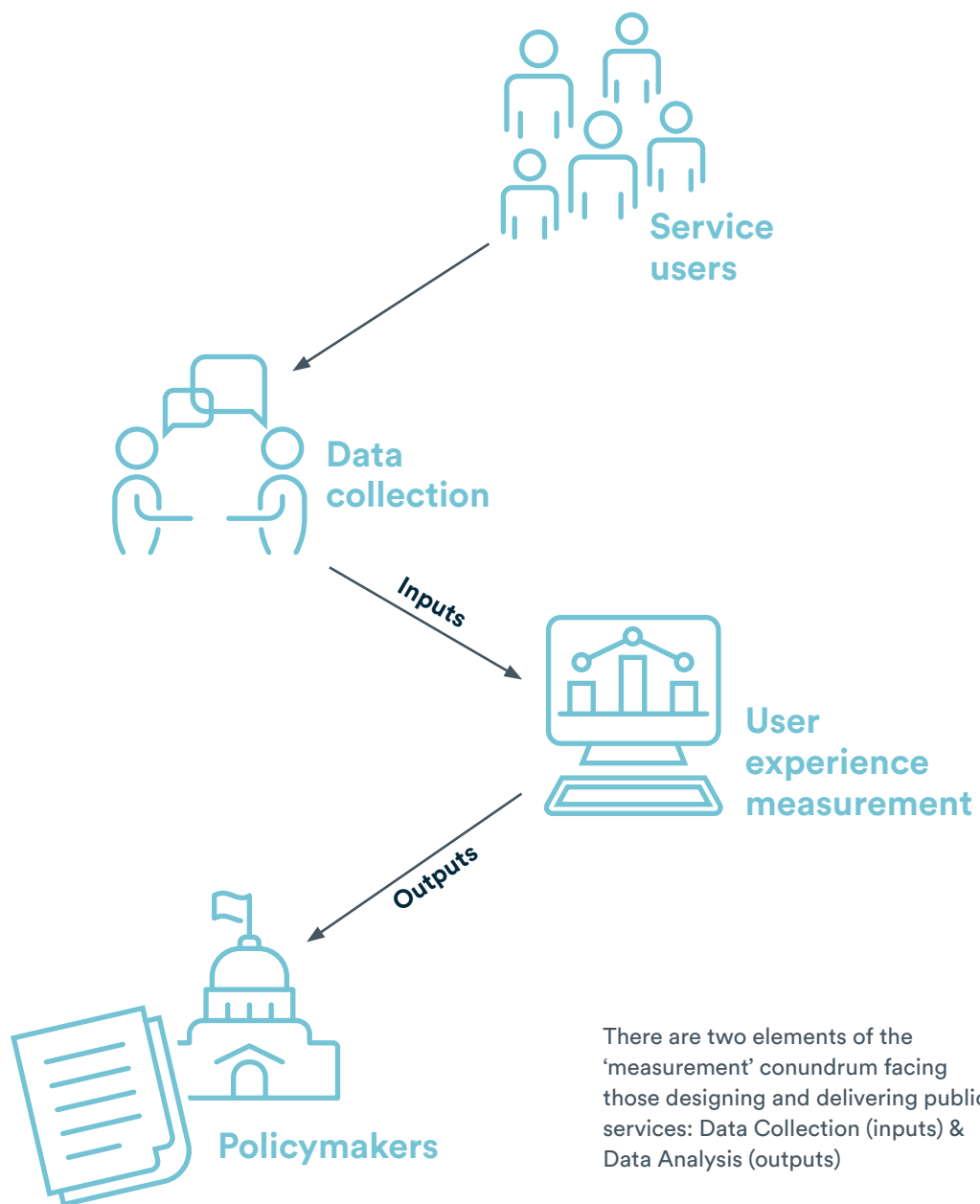
1. Data Collection (inputs):
Who and what needs to be measured, and how often and in which way is the data best collected?

2. Data Analysis (outputs):
How can the information and insights be distilled into the simplest, most comprehensible format, whilst retaining the key messages which should underpin decision-making?

Of course, these two elements are deeply intertwined – what one wants their outputs to look like will define their inputs or, conversely, what the inputs are will define the outputs. The fundamental requirement is, however, that, through a well-defined series of measurements, an accurate picture of a user's sentiments towards a service (or series of services) can be easily understood by the people who are designing and delivering them.

Service users

Data collection

Inputs

User experience measurement

Outputs

Policymakers

There are two elements of the 'measurement' conundrum facing those designing and delivering public services: Data Collection (inputs) & Data Analysis (outputs)

Data Collection – Inputs

‖ ‖

Should we measure the satisfaction of all service users, or only those with strong feelings?

Users with strong feelings are the most vocal service users, but tend to be motivated more by strong negative feelings than strong positive ones; it is therefore important that we try and develop research techniques that use a representative sample of the user base.

Deciding what to measure is only one side of the equation. Deciding whose experiences should be measured is equally important.

We don't need in-depth studies to tell us that organisations do not receive feedback from every service user. Nor can we map or track the journey of every single user. In this respect the views of some users have become more equal than others. However, for any measure of sentiment to give a good representation of how people feel about a service they need to obtain the views of a representative sample of the user population.

This can be extremely difficult to achieve, even in simple, transactional services. There are a number of studies which show that those at the more extreme ends of the spectrum when it comes to sentiment (i.e. users who are very dissatisfied and very satisfied) are more likely to make their feelings known than the average user.

This is well-illustrated by the findings of a recent survey-based piece of research of US consumers which indicated that only a fifth give feedback directly to companies (regardless of whether they had a very good or very bad experience), and that people are significantly more likely to send feedback to a company after a very bad experience than after a very good experience [41].

When it comes to government services this problem is intensified. Government service users are ever more diverse – be that linguistically, culturally or across a range of other demographic factors – and often less self-selecting than those you might find accessing other types of service. For example, a person might choose to access a commercial service, but whether they want to or not a person will be required to interact with their country's national ID system. This creates a more elaborate population of users that would need to be sampled to gain a comprehensive understanding of user satisfaction.

[41] 'How Consumers Give Feedback, 2019' – Qualtrics.

Should we use metrics or descriptive measures?

Metrics often give high-level aggregated, comparable and simple-to-understand snapshots of opinion, but without the deeper insights yielded by descriptive measures we may fail to interpret data accurately. This means a mixture of metrics and descriptive measures is optimal.

Putting aside the complexities of sampling public service users and whether any of the existing metrics are suitable for the measurement of user satisfaction with government services, some researchers have begun to question the usefulness of the prevailing user satisfaction metrics in all contexts.

Influential user experience researcher and blogger Jared Spool has argued that organisations should ignore the majority of common metrics by which businesses may compare themselves to competitors, such as the Net Promoter Score (NPS), the System Usability Score, and Customer Satisfactions scores. Spool has dismissed these 'grand unified metrics' as a myth. Instead, he has called for businesses to strive to be different from one another, and create the best value for their customers, a value which is unique to each organisation: 'success metrics should reflect the unique value proposition of each company' [42].

Fundamentally many of these metrics only offer a shallow overview of users' experiences. Furthermore, the diversity of public services, their users and the reasons for their use means that often asking someone if they are 'satisfied' – either directly or through one of the aforementioned metrics – could mean a plethora of things. This could mask real issues with user experience, or even falsely represent people's levels of satisfaction. For example, asking someone who has just been turned down for a visa through an online portal whether they would 'recommend this service to family and friends' is unlikely to glean a fair representation of the user's experience of the service, and instead capture their reaction to the policy decision regarding their application.

More descriptive measures are generally less consistent in terms of format and include larger amounts of 'free-text', making it difficult to aggregate and often time consuming to analyse.

[42] 'The Myth of the Grand Unified UX Metric' – Medium.

What balance do we strike between qualitative and quantitative measures?

Quantitative results provide easy comparisons and allow for organisations to track service performance over time, but the richer insights of qualitative data shed greater light on the drivers and motivations of user satisfaction.

This takes us on to the question of the balance between qualitative and quantitative measures.

The most familiar forms of user experience feedback are those collected through some form of scoring – from numbered scales to multi-option closed questions or even a series of pictorial representations (such as smiley faces). These are all fundamentally quantitative measures. Relatively simple to understand, collect and collate, they allow organisations to obtain and analyse a large number of responses with ease.

They are, however, relatively superficial, giving a single or small number of data points and more often than not only offering an insight as to how the user feels towards the service, not why or what made them feel that way. On the other hand, there is undoubtedly a benefit to simple quantitative measures if high volumes of feedback can be obtained which crystalise the user population's sentiment towards a service.

Tracking this data can show trends, act as a quality assurance monitor and allow for comparisons with other services, operations or alternative delivery models.

Nonetheless, quantitative data often needs to be supplemented by the collection of more long-form qualitative data. This could be collected through open ended questions or surveys, focus groups or even more in-depth methods such as 'mystery shopper' initiatives. This qualitative data gives a much deeper and richer set of insights – focusing on what drives and impacts a user's experience. However, analysing and aggregating responses is time-consuming, complex and can be reliant on the subjective interpretations of the analyst.

As such, a balance of the two types of measure can be the best route to creating an illustration how a user feels about a service and why they feel this way.

What sample size is appropriate to accurately measure user satisfaction?

Modelling and statistical methods mean that, increasingly, small sample sizes can be used to make accurate predictions about the sentiments of much larger user populations. This means researchers can prioritise data from which inferences can be effectively drawn rather than just gathering large quantities of data which cannot be used effectively.

As discussed above, qualitative data requires fewer survey subjects and produces deeper insights into the reasons why users are satisfied or dissatisfied, while quantitative results, requiring a larger sample size of participants, seek to uncover whether participants are satisfied and not why. However, an important question is: how many people should be used in a sample size when conducting research into user experience?

Many research methods use smaller groups of respondents and are able to effectively model using their sample size to make inferences about the attitudes of a much larger population: a prime example would be nationally representative polling, which is commonly used in electoral politics. Such methods of data collection have become increasingly sophisticated over time, and better able to make accurate inferences about widely held attitudes.

An alternative method of using smaller groups to measure user satisfaction would be to capture a strong, representative cross-section of the user population to examine their views: an example of this would be to use focus groups. These small groups of subjects, who share common traits or shared experiences, are brought together to examine a predetermined topic, and are a key technique used in market research. As with polling, the ways in which insights yielded by focus group participants are incorporated into market research findings have become increasingly refined over the years [43].

Statistical methods, such as MRP, can also be employed to help analyse large quantities of data to produce accurate estimations of attitudes and views – with the use of techniques such as machine learning and AI, it is likely that these methods' predictive power will continue to grow more refined over time. An important principle to follow, however, is not to harvest large amounts of data in the belief that more data will lead to more accurate predictions: what matters more in collecting data is that it is sufficiently well-defined so as to be able to be interpreted and yield good insights about a user population's sentiments. In other words, researchers should ensure that they are able to draw inferences from the data they collect rather than prioritising larger amounts of data.

The purpose of such mixed methods is to complement and validate the insights uncovered from qualitative research about the reasons for users' preferences with the digestible and actionable data points revealed by quantitative research. Qualitative research can be used to uncover themes and factors which drive satisfaction among a smaller sample of users – once no new themes emerge, researchers can begin quantitative analysis and forming insights and action points based on the information generated from qualitative research.

[43] 'Focus groups shape what we buy. But how much do they really say about us?' – Vox.

What is more revealing: stated preference or revealed preference?

Both stated and revealed preference models are vulnerable to inaccuracies owing to people's responses and the inherent irrationality of people's behaviour. However, using both may be optimal to gaining a real understanding of people's true preferences.

A well-worn debate in economics regarding stated vs revealed preference may offer some insights in relation to what basket of measures could best capture the sentiments of a service user.

'Stated preference' measures require an individual to indicate their predilections. For example, if a researcher asks an individual to rate or rank in order of preference or indicate what they believe the value of an object or service is, this would a 'stated preference' model. In these examples the person is – to point out the obvious – stating their preference.

'Revealed preference' models seek to gain insights by observing the actions of an individual, rather than asking them. In other words, a researcher tracks the actions of the individual to see what their preferences are in the belief that they will be 'revealed' by their choices. Some economists argue that this often leads to a more truthful indication of preferences. This is not necessarily just because some people lie when they are asked to rate or value something: unconsciously an individual might not realise the value they ascribe to something until the decision needs to be made.

For example, someone might state that their favourite music is the works of Mozart, whereas in fact when you track their listening habits, they seem to prefer Abba. They may have stated Mozart as they felt this made them seen more erudite and cultured, knowing that in fact they preferred Abba. On the other hand, they may have simply not realised that they actually preferred Abba.

The advantage of the 'revealed preference' method is that, by observing individuals' behaviour, researchers can infer why they are choosing to make certain decisions without asking them to state why they did so. Using 'revealed preferences' does have pitfalls, however. Firstly, it assumes people make decisions in a rational fashion – that under the same (or very similar) circumstances the preferences of an individual would remain stable. In other words, they would repeat the same set of actions, driven by the same preferences repeatedly. Of course, psychologists and behavioural economists – amongst others – have long since proved that people do not act rationally, and the repeatability and predictability of actions is significantly more complicated than tracking someone's revealed preferences.

Furthermore, a focus on revealed preferences also means that understanding the motivations of an individual will be lost. Without asking an individual to state why they took such an action you can only observe the decisions and therefore can infer why it was made.

Both approaches therefore have their critics. However, the underpinning logic of these two different methods of deciphering how to best measure people's preferences can be used to inform how we might select a range of inputs. We might want to both observe how an individual uses a service and measure their actions to reveal whether they are satisfied, whilst at the same time also asking them questions to explore why they have chosen to behave in a certain way.

Should we be asking users for their feedback, or tracking their behaviour to reveal their levels of satisfaction?

Using both active and passive feedback will paint the fullest picture of users' satisfaction levels, but measuring outcomes from interactions with government services is often more complicated than the transactional experiences of customers in commercial settings.

User experience professionals have long been using a range of data collection methods which both ask people to state their opinion and observe how they use the service. These can be broadly separated into two groups:

Active feedback

A user is asked about their opinions towards a particular service, for example, through customer satisfaction and Net Promoter Score surveys, feedback forms or focus groups. These require the user to actively indicate their levels of satisfaction. This could be through a button on a website, a verbal response to question from a telephone-handler, or the completion of a physical survey.

Passive feedback

A user's actions and outcomes are tracked in order to gauge how likely they are to be satisfied with a service. Some obvious examples include recording the time it takes to complete a task, the number of times a person returns to a service, or how many clicks on a webpage it takes for the person to find what they are looking for. Some organisations even use the number of support requests, queries to staff, and que lengths as a measure for how likely someone is to be satisfied.
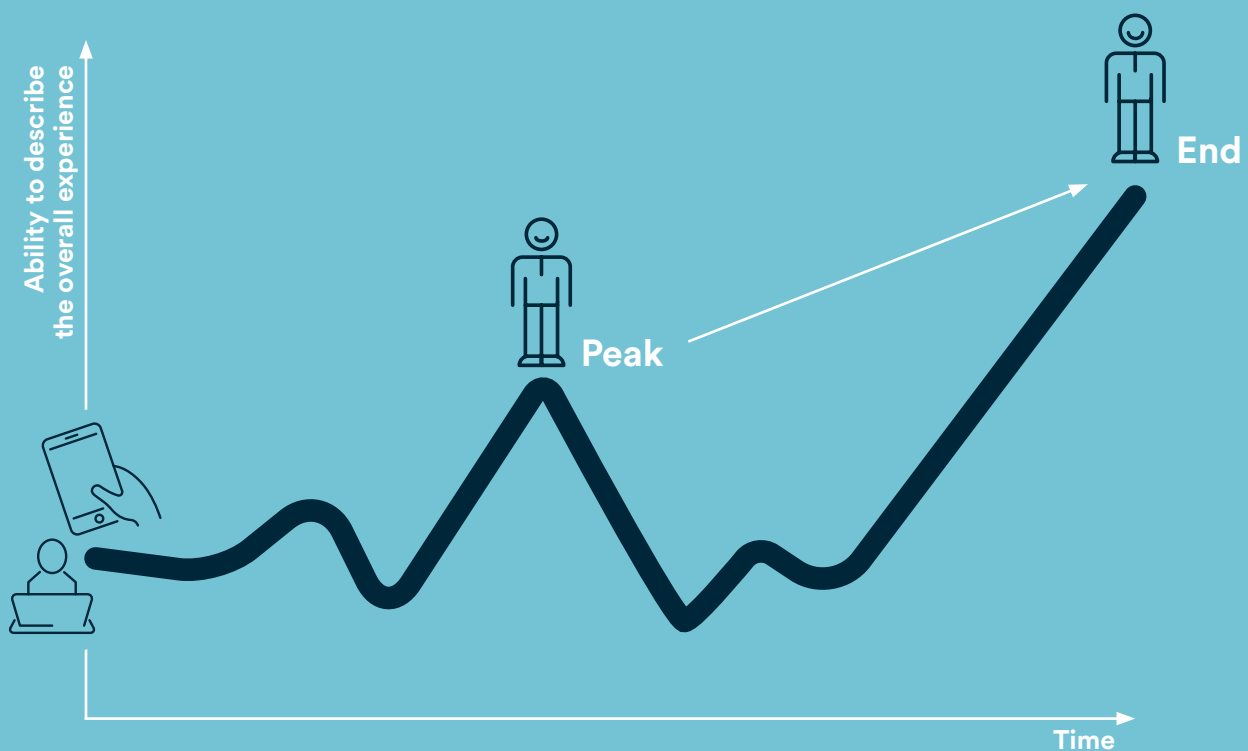
There are also increasingly innovative measures that mix active and passive feedback, such as 'social listening'. This is the tracking of the frequency and sentiments of posts on social media regarding your product or service. In many respects, it is an 'active' means of feedback as the user is – proactively – expressing a view, however, the collation of this data is in many respects passive, as organisations are simply tracking trends rather than directly collecting the feedback themselves.

These two types of feedback collection can also be thought of as a direct measure (active feedback) and proxy measure (passive feedback) of user sentiments. The former requires an organisation to seek out a user's views, whereas the latter sees organisations use metrics which it thinks represent sentiment to try and measure a user's views.

Of course, in reality a mix of both methods of feedback collection will give the fullest possible picture of users' sentiments. But the balance between the two methods of collection and how to interpret what they mean for user satisfaction are still hotly contested. Moreover, in the world of public services, where the outcomes for a service user are often more intricate than the transactional journeys more commonly found in a commercial setting, where these feedback methods are typically applied, complicates things further.

Active Feedback

Passive Feedback

Using both active and passive feedback will paint the fullest picture of users' satisfaction levels

Ability to describe the overall experience

Peak

End

Time

Research has shown, however, that users tend to decide how they feel about a service at its peak and its end, rather than the average across their whole experience.

In other words, users decide how they feel about their whole experience based on just two moments, the best or worst part of their experience, and the end of it.

This is known as the 'Peak-end Rule' and it is the best theory available as to how people will come to a conclusion on their sentiments towards a service.

At which points in a user's journey do we measure their satisfaction, and how often should we collect satisfaction data?

Researchers generally accept that measuring users' sentiments at the 'peak' and at the 'end' of their service journeys provides the most accurate picture of their satisfaction levels. However, pinpointing when exactly these touchpoints are is not simple in relation to many government services.

The who and the how has been discussed throughout the above, but so far little attention has been paid to the where and when.

Firstly, let us examine the where. This is of course contingent on the medium through which a user accesses a service; for example, in-person, online or over the phone. Each of these mediums will have different 'touchpoints' – the key junctures at which a user interacts with the service. Each of these touchpoints creates a different opportunity to collect a person's feedback. On the phone you can quiz a someone on how they found a service (active feedback/stated preference), whereas on a website you can track which series of webpages a person goes through to get to the outcome they want (passive feedback/ revealed preference).

With these 'touchpoints' in mind, we can move onto the when. It is only possible to get feedbacks at certain times – at specific touchpoints. Research has shown, however, that users tend to decide how they feel about a service at its peak and its end, rather than the average across their whole experience. In other words, users decide how they feel about their whole experience based on just two moments, the best or worst part of their experience, and the end of it.

This is known as the 'Peak-end Rule' and it is the best theory available as to how people will come to a conclusion on their sentiments towards a service. This would imply, therefore, that service designers – and those that are measuring the sentiments of users – should focus their energies on these two points in the user's journey [44].

Although it would be possible to identify the 'peak' and 'end' of a person's experience with some public services, it would not be possible to identify this cycle in all instances. Some services are iterative, or have no clear 'end'. For example, policing as a public service will only have a 'peak' or 'end' for people committing or effected by crime, it however has an important and constant role in the lives of all citizens and residents. Or, for example, waste management services would not typically be seen to have a 'peak' or an 'end' for most if not all citizens and residents, but is a crucial public service.

[44] 'Peak-end rule' – The Decision Lab.

An aggregation of inputs from users –through either active or passive feedback – these metrics are meant to create an overall picture of how a service is performing in the eyes of those that use them.
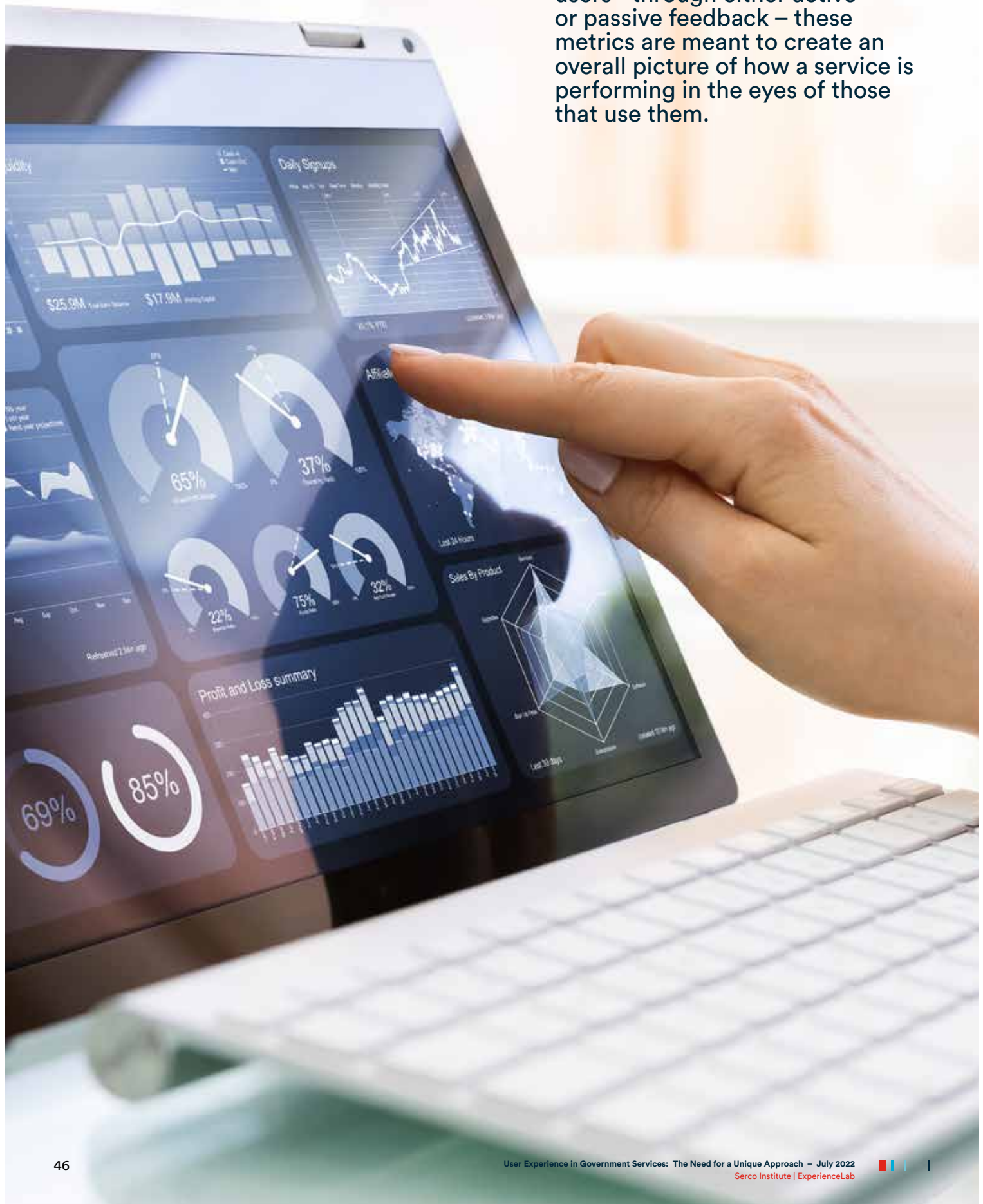
Data Analysis – Outputs

|||

What balance do we strike between high-level quantitative results and more insightful qualitative data?

Qualitative data offers a more accurate understanding of why service users feel the way they do and is essential to understanding why users behave as they do, but given the very large user-base of government services, quantitative data is essential to collating the insights yielded by qualitative results and to ensuring they are representative of the user base. In practice, therefore, results will have to rely at least in part on quantitative results.

A small number of metrics is how we most often see measures of customer satisfaction articulated. An aggregation of inputs from users –through either active or passive feedback – these metrics are meant to create an overall picture of how a service is performing in the eyes of those that use them. They are simple to comprehend and are meant to easily allow for the comparison of two or more services by quantifying the sentiments of multiple users into a single comparable figure.

As discussed, these figures are, however, a straightforward expression of overall sentiment, and rarely offer insight into the reasons why people feel a particular way about a service. Qualitative data offers a greater insight into why a person feels a particular way. Reading individual qualitative responses can take time and requires interpretation. Furthermore, to gain an illustration of the sentiments of a representative sample of the user-population a significant number of qualitative responses will need to be collected and analysed. Therefore, these qualitative responses can themselves often be translated into quantitative metrics by the time they become an 'output' that will be read by policymakers. This can, however, remove a significant amount of the 'richness' of the insights.

Public services often have a very large user-base, with a diverse set of needs. They are also Likely to access services through a range of different mediums – in-person, over the phone, digitally etc. This means the collection of both aggregable, comparable quantitative and qualitative feedback is difficult. Furthermore, policymakers have a range of different issues beyond user experience that needs to be considered – for example, the accessibility of the service for a range of users and legal requirements. Equally, policymakers will rarely be experts in user experience and – with a range of considerations already requiring focus – may not take the time to properly consider any outputs of user experience feedback if they are not quickly and easily understandable. The balance between creating insight-rich as well as easily discernible and usable analysis of user feedback is particularly difficult and critical when it comes to public services.

Should we rely on descriptive statistics or inferential statistics?

Inferential statistics, which allow researchers to make predictions about a large number of users based on a smaller sample size, yield the greatest insights into what drives satisfaction with government services, and efforts should be made to move towards generating more inferential results.

When it comes to the statistical output of service-user sentiment measurement, there are broadly two types of figures that can be produced.

Descriptive statistics focus on outlining the visible characteristics of a dataset (a population or sample). Meanwhile, inferential statistics focus on making predictions or generalizations about a larger dataset/population, based on a sample.

As their name implies, descriptive statistics seek to 'describe' the character of a sample. These types of statistics offer insights into the features of a group of people, such as how often a characteristic appears, the scale of different characteristics and the averages of characteristic within a population. For example, an average customer satisfaction score of all responses received in an online end of service survey is a descriptive statistic. Equally, a figure outlining the frequency by which women or men responded to the survey would be a descriptive statistic. In effect, all of these types of statistics summarize the key features of the data being analysed.

Inferential statistics on the other hand make predictions based on a sample about a larger population. These types of statistics also allow analysts to estimate and predict future outcomes. Therefore, its results are usually expressed in the form of a probability. For example, how likely a characteristic identified in a sample is representative of an entire population. Or to put it another way, how like the responses to a survey are to be representative of the entire user-population.

User experience is typically expressed using descriptive statistics. Let us take, for example, the sentence 'four in ten users we surveyed said that they were happy with their experience'. This is the simplest way to articulate feedback, but is unlikely to be representative of the entire user-population. As discussed above, capturing a representative sample of the service-user-population is difficult as users who have a particularly positive or negative experience are more likely to give feedback. If efforts can be made, however, to collect a range of feedback which closely resembles the user-population, inferential statistics can be used to generalise responses and understand the view of the wider user-population, as well as predict their future behaviour. This allows policymakers and service designers to create services that meet the wants and needs of those that use them.

As noted, it is difficult to collect feedback which effectively reflects the service-user (and potential service-user) population. This is particularly difficult due to the complexities of who uses public services. Therefore, creating user feedback measurement outputs which include inferential statistics can be counter-productive when it comes to public services. If an inaccurate sample of feedback is collected, incorrect inferences could be created or analysis will include large margins of error. As such, it would be irresponsible to create policy predicated on such analysis.

That said, inferential statistics will offer the greatest insights into how a large number of users feel about a service. It is therefore critical that efforts are made to increasingly move towards a system whereby inferential statistics can be generated and used by policymakers to understand and improve user experience in public services.

How can we design satisfaction metrics to allow us to track changes in sentiment over time?

Tracking changes in sentiment over time is relatively simple: researchers can put the same questions about the same services to the public repeatedly over time and monitor how levels of satisfaction change. However, changing external circumstances unrelated to the quality of public services may influence how users report satisfaction.

User satisfaction metrics are often designed to allow organisations to compare their 'scores' over time. This offers a useful means by which they can rank performance, track changes in satisfaction and identify areas for improvement.

This is relatively simple to do, by asking similarly sized samples of service users the same set of questions about their sentiments towards government services over a protracted period of time. With each 'wave' of polling, changes in levels of satisfaction can be tracked, and so researchers can monitor how public satisfaction with government services may be rising or declining with time. This is, in effect, what the Serco Institute is doing in a separate research project: People Powered Public Services, in which the Institute monitors changes in sentiment towards a wide range of public services through rounds of polling conducted over time in the UK and in Australia.

However, it is important to note that reported changes in sentiment cannot necessarily be attributed solely to public satisfaction with the services in question or to real improvements or declines in the quality of services provided. A number of changing contextual factors influence how people may perceive services from month to month. These may be out of the control of policymakers, but nevertheless impact upon users' reported levels of satisfaction. Furthermore, respondents may not distinguish between their feelings towards the public service itself and their feeling towards the associated policy area, making it harder to measure satisfaction. For instance, an unpopular public transport policy introduced in March may result in a significant decline in reported satisfaction with public transport services between January and May, as users do not make a distinction between their satisfaction with transport policy and their satisfaction with the actual transport services themselves.

How can we design satisfaction metrics to allow us to compare results to other public services?

The sheer diversity of services offered by governments means that designing a single universal metric for satisfaction with all government services is especially difficult.

User satisfaction metrics are often designed to allow organisations to compare The diversity of services on offer – and even the diversity of how even a single service is offered/accessed – makes creating a comparable series of metrics across public services a challenge. For example, a metric seeking to measure how likely a user is to recommend a service might not be appropriate if the service they are using is not one they would 'choose' to use, such as an ambulance service or tax payment system. Equally, many passive measures, such as those seeking to measure user retention or number of transactions made may not be appropriate across all services.

Furthermore, as noted in other sections, certain policy areas evoke strong emotional reaction among users, and as such government services in these policy areas are less likely to be evaluated by users purely on the merits of the delivery and quality of services. To take an earlier example, a user of a government immigration service who has just had their visa application rejected is unlikely to provide an objective assessment of their experience using the service. In other cases, unpopular public policies may lead users to negatively rate the services associated with those policies without drawing a distinction between the service itself and the policy.

A unified, universal metric (or small series of metrics) across public services is therefore difficult to identify from the current range of options. This makes creating outputs which are easily comparable difficult.
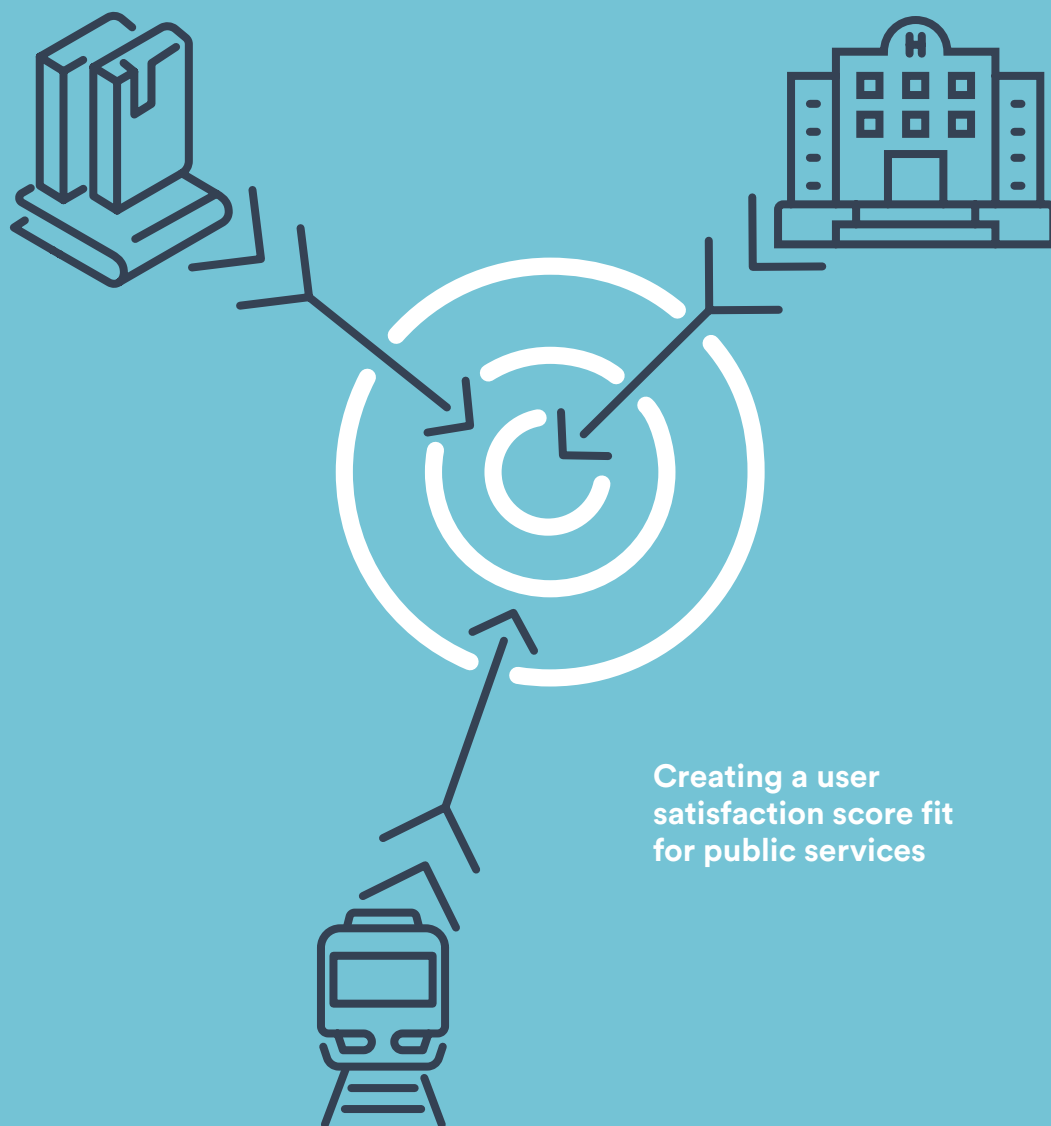
Creating a user satisfaction score fit for public services

Creating a user satisfaction score fit for public services – Hitting a moving target

Undoubtedly many governments are making significant efforts to increase their capacity and improve their capability to measure user satisfaction with public services.

Putting aside the need to keep up with the latest research and techniques, too often governments are reliant on metrics and measures which were not designed for deployment in the environments which they are being used. As such it is critical that – alongside the need for new techniques and process to be better embedded – specially designed metrics and methods of measurement are developed so the correct inputs can be gathered and outputs developed to help policymakers deliver ever more user-centred public services.

These will need to be predicated on a mixture of active and passive feedback, gleaning users' stated and revealed preferences across the key touchpoints of services, through quantitative and qualitative inputs and outputs, and expressed to policymakers using both descriptive and increasingly inferential statistics.
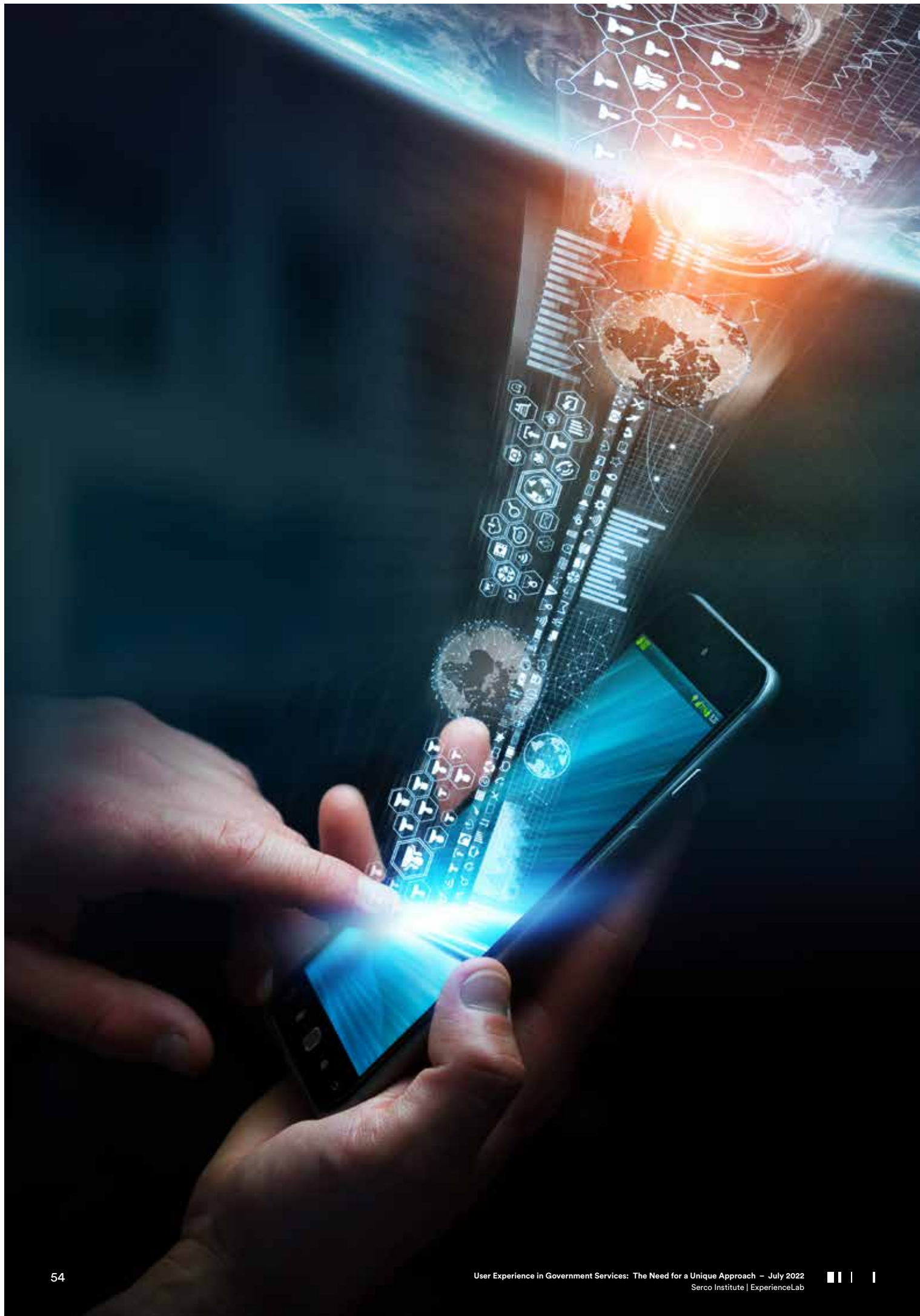
▮▮ ▏ ▮ Conclusion – The search for a new measure

1.   Where next?

Where next?

This short paper has attempted to outline some of the current approaches to measuring satisfaction with government services, the need to re-evaluate those measures and the scale of the challenge.

Although there are examples of good practice across the globe, there is a real opportunity for governments to take up the challenge of better understanding the wants, needs and feelings of the people that use their services. The dividends for doing so will not only be better public services, but happier, more trusting citizens and residents.

This report is fundamentally an articulation of the problem statement: we have outlined existing practice around the world and attempted to identify the main, often unique, challenges in accurately measuring satisfaction with government services. The next step is to offer up some potential solutions. Working with the user experience experts at ExperienceLab, in future papers we will explore some alternative approaches and offer up some new measures which we believe could better capture the views of public service users.

We would welcome input from readers and other stakeholders in user experience research, and would encourage interested parties to submit additional questions or feedback pertaining to user experience and the content of this report via: info@sercoinsitute.com .

This will guide future work and papers we conduct on this subject matter.