Mobile Connectivity Index

Methodology

August 2019

GSMA

The GSMA represents the interests of mobile operators worldwide, uniting more than 750 operators with nearly 400 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces the industry-leading MWC events held annually in Barcelona, Los Angeles and Shanghai, as well as the Mobile 360 Series of regional conferences.

For more information, please visit the GSMA corporate website at www.gsma.com.

Follow the GSMA on Twitter: @GSMA

GSMA Connected Society

The Connected Society programme works with the mobile industry, technology companies, the development community and governments to increase access to and adoption of mobile internet, focusing on underserved population groups in developing markets. Key activities include:

- **Generating and disseminating insights and learnings** on the mobile internet coverage and usage gap;
- **Supporting mobile operators** to extend coverage and drive usage; and
- **Undertaking advocacy and policy work** to ensure that mobile operators' efforts to achieve greater digital inclusion are being effectively supported by governments, the international community and other stakeholders.

For more information, please visit www.gsma.com/connected-society

GSMA Intelligence

GSMA Intelligence is the definitive source of global mobile operator data, analysis and forecasts, and publisher of authoritative industry reports and research. Our data covers every operator group, network and MVNO in every country worldwide — from Afghanistan to Zimbabwe. It is the most accurate and complete set of industry metrics available, comprising tens of millions of individual data points, updated daily.

GSMA Intelligence is relied on by leading operators, vendors, regulators, financial institutions and third-party industry players, to support strategic decision-making and long-term investment planning. The data is used as an industry reference point and is frequently cited by the media and by the industry itself.

Our team of analysts and experts produce regular thought-leading research reports across a range of industry topics.

www.gsmaintelligence.com

info@gsmaintelligence.com

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.

This document has been financed by the Swedish International Development Cooperation Agency, Sida. Sida does not necessarily share the views expressed in this material. Responsibility for its contents rests entirely with the author.

Contents

1. Introduction

About the Mobile Connectivity Index

In 2018, almost 300 million people connected to mobile internet for the first time, bringing the total connected population to more than 3.54 billion people globally (or 47% of the global population).[1] For many of these individuals, mobile is the only method of accessing the internet, so growth in mobile internet adoption also drives digital inclusion, especially in low- and middle-income countries (LMICs).[2]

Mobile internet connectivity is not however equitable; just over 40% of the LMIC population (around 2.6 billion people) are connected, compared to almost 75% of the population in high-income countries. More than half the world's population are therefore still unable to realise the social and economic benefits that mobile internet can enable. If current trends continue, more than 40% of the population in LMICs will still be offline in 2025.

To support the mobile industry's efforts to drive mobile internet connectivity and accelerate digital inclusion, the Mobile Connectivity Index (MCI) measures and tracks enablers of mobile internet connectivity. The Index has been constructed according to the steps set out in the guidelines developed by the OECD and the European Commission's Joint Research Centre (JRC).[3] This methodology for the MCI presents the theoretical framework that underpins the Index; the process for selecting the indicators, along with how they are structured; the approach used to normalise the data; the weights used in the Index; and the approach to aggregation.

Theoretical Framework

What is measured?

The Index measures the enablers of mobile internet connectivity. It is therefore an input index. An input index measures a number of indicators that lead to an important outcome, in this case why people are (or are not) using mobile internet connectivity. This is different to an output index which, in the context of mobile connectivity, might seek to measure and understand how (or how much) people are using mobile internet services.[4]

[1] Source: GSMA Intelligence
[2] This categorisation is based on the World Bank's country classification and includes countries that the World Bank classifies as low-income, lower-middle income and upper-middle income.
[3] Handbook on constructing composite indicators: methodology and user guide, OECD and JRC, 2008
[4] In order to track outcomes and the impact of mobile, the GSMA publishes the annual Mobile Industry SDG Impact report.

Why is an index necessary?

There is no single barrier or enabler to mobile internet connectivity; rather, a number of prerequisites are necessary for a country's population to use mobile internet services. An index is required because it measures multiple enablers and can summarise the complex reality of delivering mobile internet connectivity in any country.

A number of indices exist in the ICT sector, including:

- ICT Development Index (International Telecommunication Union);
- Affordability Index (Alliance for Affordable Internet);
- Digital Economy and Society Index (European Union);
- Global Connectivity Index (Huawei); and
- Inclusive Internet Index (Facebook and the Economist Intelligence Unit).

The MCI has been designed to ensure that it does not replicate any of these or other related indices. In this respect, the index has four key characteristics that together distinguish it from other indices:

- It focuses specifically on mobile internet connectivity rather than internet connectivity in general (including fixed-line connectivity). Given that the internet gap in low- and middle-income countries (LMICs) is expected to be addressed to a significant extent by mobile, it is important to understand the enablers of mobile connectivity specifically.
- As the index is focused on mobile internet connectivity, the majority of the underlying indicators are unique to the MCI and are either not available or not used in other indices (for example, spectrum availability, taxation on mobile services and the availability of content on mobile platforms).
- It is an input index that seeks to measure the performance of countries against a set of key enabling characteristics, rather than an output index that measures internet take-up and usage.
- It is a global index, encompassing 165 countries that account for more than 99% of the world's population.

The MCI scores are highly correlated with mobile internet adoption, while countries that achieve significant improvements on the MCI are also more likely to have seen increases in mobile internet adoption over time.[5] This means that the MCI is an effective tool to identify priorities to drive mobile internet adoption.

How are the enablers measured?

The enablers of mobile internet connectivity that inform the indicators selected for the Index are:

1. **Infrastructure:** the availability of high-performance mobile internet network coverage;

2. **Affordability:** the availability of mobile services and devices at price points that reflect the level of income across a national population;

3. **Consumer readiness:** citizens with the awareness and skills needed to value and use the internet; and

4. **Content and Services:** the availability of secure online content and services accessible and relevant to the local population.

[5] See State of Mobile Internet Connectivity 2018, GSMA

2. Data selection

Selection criteria

As the Mobile Connectivity Index is an input index, it is important that each indicator is an 'input' for mobile internet connectivity rather than an output or outcome (e.g. measuring the level of take-up). It is also important to develop a set of criteria against which each indicator can be considered for inclusion in the Index. The following criteria have therefore been developed, based on guidance from the JRC and OECD:

- **Relevance**: the indicator should measure a barrier or an enabler in the take-up of mobile internet services;
- **Accuracy**: the indicator should correctly estimate or describe the quantities or characteristics they are designed to measure;
- **Coverage**: the data should cover as many countries as possible, as the Index is intended to be a global index. An indicator is not included if there is missing data on more than 25% of countries in the Index.
- **Timeliness**: the data should be collected consistently over time.

A key consideration in the assessment of accuracy is to include, to the greatest extent possible, '*hard*' indicators that are objective and can be quantified. These are distinct from '*soft*' indicators that are usually based on qualitative data from surveys or case studies. Such indicators are typically used to measure things that are difficult to quantify – for example, the quality of governance or the extent of corruption. Although soft indicators are very useful for some indices, particularly those where hard indicators are difficult to develop, they are not used in the MCI. This is to ensure that operators, regulators and other stakeholders have objective benchmarks on which to target improved performance.

Although the indicators included in the Mobile Connectivity Index have all been carefully chosen based on the above criteria, there are some cases where data constraints require the use of proxy indicators if it is not possible to measure an enabler with complete accuracy. For example:

- Indicators such as international bandwidth per user, number of secure servers and number of Internet Exchange Points (IXPs) are included as proxies for the quality of a country's backhaul and core network. This is because high quality end-to-end mobile services require a resilient and high-capacity backhaul and core network.

- There is currently no data comparing a large number of countries in the area of digital skills or awareness. More traditional skills indicators are therefore used to measure consumers' ability to effectively use and engage with digital technology (for example, literacy and years of schooling).

- Indicators measuring the availability of mobile-specific content primarily focus on smartphone applications due to the lack of comparable data across countries on other types of mobile content.

Although the vast majority of the indicators are highly correlated with mobile internet penetration, suggesting that on average they are associated with higher take-up, there may be specific countries where they work less well as proxy indicators. These indicators will therefore be reassessed going forward and, where they can be improved, incorporated into future versions of the MCI.

Indicators

Table 1 details the indicators that make up the Index. The Index comprises four key enablers, which in turn comprise a number of dimensions. These dimensions are constructed by aggregating one or more indicators.

Table 1: Mobile Connectivity Index Indicators

Enabler	Dimension	Indicator	Original unit of measurement	Source
Infrastructure	Network Coverage	2G network coverage	% of population covered	ITU
		3G network coverage	% of population covered	GSMA Intelligence
		4G network coverage	% of population covered	GSMA Intelligence
	Network performance	Mobile download speeds	Mbps	Ookla's Speedtest Intelligence
		Mobile upload speeds	Mbps	Ookla's Speedtest Intelligence
		Latencies	Milliseconds	Ookla's Speedtest Intelligence
	Other enabling infrastructure	International bandwidth per user	Bits per second	ITU
		Number of secure servers	Secure servers per 1 million people	World Bank
		Access to electricity	% of population with access	World Bank
		Number of Internet exchange points (IXPs)	IXPs per 10 million people	Packet Clearing House

	Spectrum	Digital dividend spectrum (a)	MHz per operator (b)	GSMA Intelligence
		Other spectrum below 1GHz	MHz per operator (b)	GSMA Intelligence
		Spectrum in bands 1-3GHz	MHz per operator (b)	GSMA Intelligence
Affordability	Mobile tariffs	Cost of 100MB	% of monthly GDP per capita	Tarifica
		Cost of 500MB	% of monthly GDP per capita	Tarifica and ITU
		Cost of 1GB	% of monthly GDP per capita	Tarifica
	Handset price	Cost of entry-level internet-enabled handset	% of monthly GDP per capita	Tarifica
	Inequality	Inequality in income, Atkinson measure	Index value (0-100)	UN
	Taxation	Cost of taxation	Cost of tax as a % of TCMO (c)	GSMA Intelligence
		Cost of mobile-specific taxation	Cost of mobile-specific taxes as a % of TCMO	GSMA Intelligence
Consumer Readiness	Basic skills	Adult literacy rate	% of literature adult population (above 15 years old)	UN
		School life expectancy (d)	Years	UN
		Mean years of schooling (e)	Years	UN
		Tertiary enrolment rate	%	UN
	Gender equality (f)	Gender years of schooling ratio	Female/male ratio	UN
		Gender account access ratio	Female/male ratio	World Bank Global Findex
		Gender Gross National Income (GNI) per capita ratio	Female/male ratio	UN
		Women, Business and the Law Index	Index value (0-100)	WBL
		Gender gap ratio for social media use	Female/male ratio	We Are Social and Facebook Audience Insights
		Gender gap ratio for mobile phone ownership	Female/male ratio	GSMA Intelligence and Gallup World Poll
	Mobile ownership	Penetration of mobile users	% of population	GSMA Intelligence

Content and Services	Local relevance	Top-Level Domains (TLDs) per capita (g)	Number of domains per person	ZookNIC
		E-Government services (h)	Index value (0-1)	UN
		Mobile social media penetration	% of population	We Are Social
		Mobile application development	Number of active mobile apps developed per person	Appfigures
	Availability	Number of apps in national language	Number of mobile apps available in national language(s)	Appfigures, Ethnologue and GSMA Intelligence
		Accessibility of top ranked apps	Average of the % of population that can use each app in the top 400 for the country	Appfigures, Ethnologue and GSMA Intelligence
	Online security	Global Cybersecurity Index	Index value (0-1)	ITU

(a) Digital dividend spectrum refers to spectrum in 600, MHz, 700 MHz and 800 MHz bands that are particularly well suited to achieving wider coverage.
(b) When constructing the metric on spectrum per operator, we exclude operators with very small spectrum holdings and market shares (e.g. operators only active in specific regions or that provide niche services).
(c) 'TCMO' refers to the 'Total cost of mobile ownership', which is the total cost to a consumer of owning and using a mobile phone, expressed in monthly terms. TCMO sums three cost categories: the handset price (amortized over an assumed handset life of three years for low- and middle-income countries and two years for high-income countries)[6]; the activation and connection price; and the price related to the use of voice, SMS and/or data.
(d) This is the total number of years of schooling (primary to tertiary) that a child can expect to receive given current enrolment rates. It is therefore a forward-looking indicator.
(e) This measures the average number of years of education received by people aged 25 and older, based on current attainment levels. It is different from school life expectancy because the latter is calculated using enrolment rates.
(f) Each of the indicators in this dimension (with the exception of the Women, Business and Law (WBL) Index) are calculated by dividing the relevant female indicator (e.g. female GNI per capita) by the relevant male indicator (e.g. male GNI per capita).
(g) This includes the number of generic top-level domains (gTLDs) registered in a country and the number of registered country-code top-level domains (ccTLDs).
(h) This indicator uses the Online Service Index score in the UN's E-Government Survey.

[6] See GSMA, Rethinking taxation to improve mobile connectivity (2019) for further details on the TCMO framework.

In this year's MCI, the following indicators were added:

- Within the **Consumer Readiness enabler**, a dimension on mobile ownership was added that includes an indicator measuring unique mobile subscriber penetration.[7] Owning and using a mobile phone is a pre-requisite to being able to use mobile internet services and existing mobile users are likely to be more aware and have the relevant skills to use mobile internet than non-users. Including this indicator/dimension means we are now measuring a key enabler that was not previously captured.

- Within the **Gender Equality dimension**, the following changes were made to better reflect the drivers of the gender gap in mobile internet use[8]:

    - Removal of gender parity ratio for literacy, as there is already a metric for education (gender parity in mean years of schooling);

    - Removal of gender parity ratio for labour force participation, as it was very highly correlated with the gender parity ratio for GNI per capita (with a correlation coefficient greater than 0.9);

    - Addition of country score for the Women, Business and the Law (WBL) Index[9] – this measures whether laws and regulations restrict women's economic and social opportunities;

    - Addition of the gender gap in social media use[10] – this captures several elements, including a gender gap in local content as well use of technology and social norms; and

    - Addition of the gender gap in mobile phone ownership[11] – this also captures the gender gap in use of technology and social norms, as well as measuring the gender gap in phone ownership.

[7] Unique subscriber penetration refers to the number of unique users who have subscribed to mobile services (excluding M2M) mobile services. Subscribers differ from connections such that a unique user can have multiple connections. Unique subscriber data is sourced from GSMA Intelligence, combining data reported by mobile operators with the annual GSMA Intelligence Consumer Survey.
[8] For further discussion on this, see GSMA, Gender Gap Report 2019
[9] Data source: World Bank Women, Business and Law Index.
[10] Data source: Facebook Audience Insights and We Are Social.
[11] Data source: GSMA Intelligence and Gallup World Poll

- Lastly, within the **Content and Services enabler** a new dimension was added for Online Security, which incorporates the ITU's Global Cybersecurity Index as an indicator.[12] This was included as safety and security is an increasingly important enabler of mobile internet adoption and usage. The GSMA will review this dimension going forward and incorporate additional indicators on safety and security where they meet our criteria for including an indicator in the Index.

For each of these indicators, data has been incorporated in all years of the Index (2014 to 2018) – including revising previous indexes – to ensure that year-on-year changes are comparable. As a result, Index and Enabler scores may differ to those published in previous years.

Pricing data

Pricing data is provided by Tarifica, with retail prices captured as of the first quarter of 2017, 2018 and 2019, including all relevant taxes.

Mobile tariffs

In order to produce comparable price metrics across countries, three baskets were defined based on usage allowance, contract and technology. The baskets were designed to capture entry or basic usage as well as more intense users. In order to construct the baskets, the following information was taken into account:

- Historic trends in average data consumption across countries, sourced from GSMA Intelligence, Ofcom, Tefficient and Opera. Future data requirements (which are likely to increase) were taken into account. We also gave due consideration to the fact that average values are often distorted by particularly intensive users of mobile services.

- A selection of allowances currently offered by operators in developed and emerging markets, provided by Tarifica.

- Baskets used in other mobile pricing benchmark studies published by the OECD, Tarifica, the United Kingdom's telecommunications regulator Ofcom, European Commission and the ITU. These represent basket designs often used in economics literature that analyse pricing in the mobile industry.

The baskets resulting in from this analysis are summarised in **Table 2**.

[12] For further details, see https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx

Table 2: Usage basket profiles

Basket	Entry	Medium	High
Monthly usage allowance	100 MB data	500 MB data	1 GB data
Tariff	Prepaid	Any	Any
Technology	2G, 3G or 4G	3G or 4G	3G or 4G

Having defined these baskets, Tarifica researched all tariffs offered by operators in each country and selected the cheapest available plan under which the basket requirements could be met. In addition, the following guidelines were applied to ensure prices were representative of regular usage and consumption patterns:

- Prepaid plans lasting less than one month were included – in such cases, the usage allowance and price were scaled up to one month to ensure comparability across tariffs (e.g. the usage and price of a five-day plan were multiplied by 6 to derive a monthly usage and price);
- Short-term promotional offers were excluded;
- Plans targeted or restricted to certain profiles (e.g. youth, student, senior) were not included; and
- Where a tariff included an initial one-off fee (e.g. activation, SIM card), this was amortised over a period of 24 months.

Prices were captured in local currencies for all countries. These were then converted into US dollars using exchange rates as of Q1 2017, Q1 2018 and Q1 2019. This approach was used to obtain pricing data for 2016, 2017 and 2018 respectively. To derive pricing data for 2014 and 2015, we incorporated data from the ITU, which has historically collected data on the 500MB basket.[13] Using ITU pricing data for 2014-2016, we calculated growth rates between 2015-2016 and 2014-2015. These growth rates were applied to the 2016 prices obtained from Tarifica to give a complete historic dataset for each basket.

[13] See the ITU Measuring Information Society Reports. The ITU's approach to collecting pricing data differs to the approach used by Tarifica, particularly because the ITU only collects data from the largest mobile operator whereas Tarifica considers all operators.

Handset price

As the MCI is particularly focused on connecting the unconnected, we are interested in measuring prices of entry-level handsets that allow users access to the internet rather than high-end devices that are sold at premium prices. In order to obtain this data, Tarifica researched the cheapest handset available in each market with internet-browsing capability. This device could therefore be a smartphone[14] or a feature phone[15]. Given that the performance for basic internet mobile applications (such as basic video or social networking) is only functional with 3G and 4G, this analysis excluded devices with 2G and WAP connectivity. Device prices were collected through researching the devices available on the websites of all MNOs in each country; other retailers' websites were analysed for the countries where MNOs did not offer handsets. This approach was taken due to the significant resource that would be required in order to inspect all non-MNO retailers in each country. However, this approach means that in some markets there may be cheaper devices available – but by keeping the approach consistent across countries, the relative differences across countries should be similar and minimise the risk of bias.

As with mobile tariffs, prices were captured in local currencies. These were then converted into US dollars using exchange rates as of Q1 2017, Q1 2018 and Q1 2019. This approach was used to obtain pricing data for 2016-2018. To derive pricing data for 2014 and 2015, we used data from Strategy Analytics on the average selling price (ASP) of handsets for each country and estimated historic prices by applying ASP growth rates to 2016 data.

Taxation

The taxation indicators are developed by estimating the proportion of the total cost of mobile ownership (TCMO) that are: (i) accounted for by all taxes and (ii) accounted for by mobile-specific taxes.

The TCMO is calculated in monthly terms on the basis of three building blocks:

- **Handset price:** representing a one-off cost that can be spread over the lifecycle of the device (after which it is assumed to be replaced). Handset prices were converted to a monthly price based on a handset lifecycle of three years for developing markets and two years for developed markets, in order to take into account differences in usage patterns, disposable income and willingness to pay.[16]

- **The activation and connection price or any other charges incurred to connect to the MNO's network**. For prepaid customers, this usually consists of an initial charge for

[14] A smartphone is a device that has an open operating platform (where new applications can be developed and installed by the user).
[15] A feature phone is a device with a closed platform.
[16] See GSMA, Rethinking taxation to improve mobile connectivity (2019) for further details on the TCMO framework.

activating the SIM card. For postpaid customers, there may be additional upfront costs, such as an initial charge for activating the number. Activation and connection prices were converted into monthly prices assuming they follow the lifetime of the device.

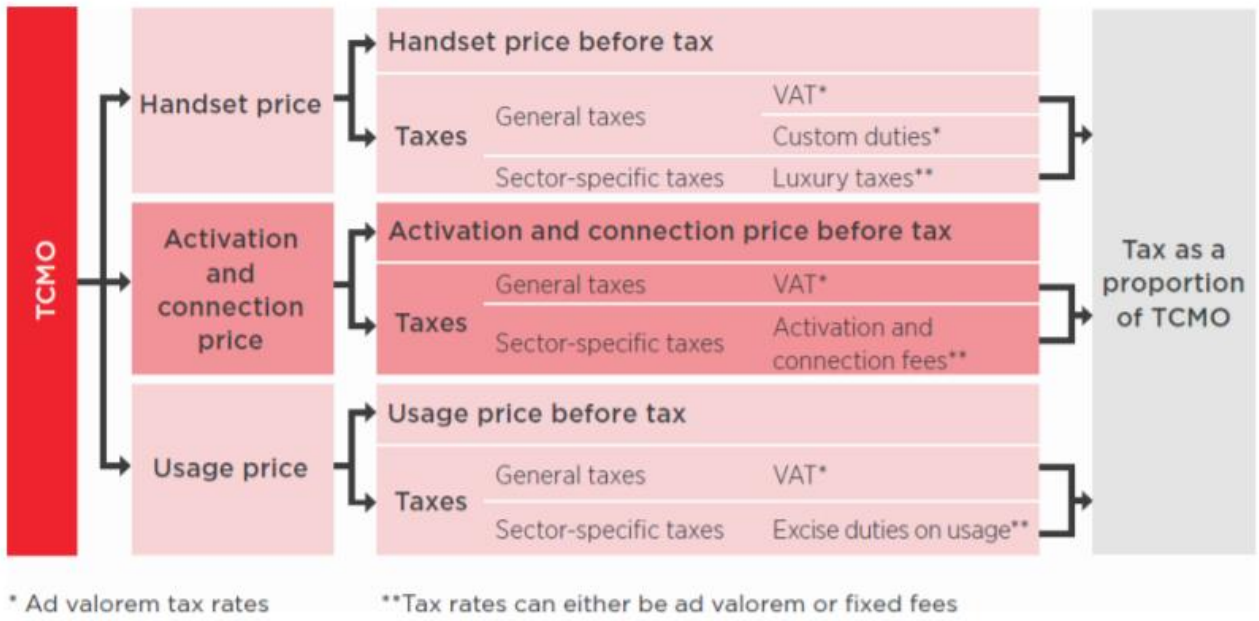
- **The price related to use:** comprising of voice, SMS and data charges. This is already expressed in monthly terms.

The TCMO was calculated for each basket *b* of country *i* as follows:

$$TCMO_{bi} = \frac{\text{Handset price}_{i}}{\text{Handset lifecycle}_{i}} + \frac{\text{Activation and connection price}_{bi}}{\text{Handset lifecycle}_{i}} + \text{Usage price}_{bi}$$

In order to calculate tax as a proportion of TCMO, taxes in Figure 1 were considered.

Figure 1: Calculation of proportion of tax in TCMO



Taxes in the TCMO were calculated by applying tax rates to the appropriate tax base.

- In the case of ad valorem tax rates (VAT and excise duties), the relevant tax base is the retail price of the relevant TCMO building block that was used.

- In the case of customs duties, the selected tax base was the retail price of the device building block in the TCMO.

- In the case of fixed amount taxes, activation and connection fees were applied on the value of the SIM card. For general fixed fees, tax payments were converted to a monthly level.

Estimates of the proportion of TCMO accounted for by all taxes and mobile-specific taxes were derived for all baskets (entry, medium and high). Subsequent analysis showed a very high correlation (above 0.95) between baskets and so only the entry-level basket has been used in the Index to compile the taxation indicators.

Availability of content

The Content and Services enabler consists of three dimensions: local relevance, availability and online security. 'Local relevance' measures the amount of content produced in a given country, including e-government services, web domains, social media[17] and mobile applications. These are included because content that is created or developed within a country is likely to be relevant to many of the people living there.

However, many people consume content that is produced outside of their own country, so it is important to measure the extent to which this more widely available content is accessible and relevant to users. We therefore developed two indicators to assess whether a country's population has content that they can understand and with which they can engage. We did this by using language as a measure for accessibility and relevance – if an individual has content available in a language they speak then they are more likely to find the internet useful.

Each of the indicators required data on the languages spoken in each country, which is sourced from Ethnologue. Specifically, a dataset was provided that had a list of languages spoken in each country as well as data on the proportion of population speaking each language, both as a first language and in total.

The second data source is a list of all mobile applications available on Google Play, Apple Store, Windows, Amazon and other smaller platforms. This dataset was provided by Appfigures. For each application, information is provided on the languages and countries it is available in as well as the

[17] This provides a platform to generate content that local populations need or are interested in.

application category (gaming, education, health etc), the year of release and whether or not it is active. Data was also provided on the top 400 ranked applications (based on downloads) for Google Play and Apple Store at the end of each year since 2014, giving an indication of the content most popular in each country.

Using these data sources, we developed two indicators that assess the extent to which a country's population has content they can understand and engage with by mapping the languages spoken in each country against the language of mobile applications.

Number of applications accessible in the national language(s)

This indicator measures the number of active applications available to a country's population in the national language. If a country has more than one national language (e.g. India and South Africa) then we include any application that is available in at least one of the country's national languages. Where a significant proportion of the population do not speak a national language (for example many African countries have French or English as one of their national languages but it is not spoken by everyone in the country), we adjust the calculation based on the proportion that speak each language. For example, if only 20 per cent of people in a country speak English, we calculate the score for the indicator and multiply it by 0.2 to account for the fact that 80 per cent of the population cannot access mobile content in English. We also apply a logarithmic transformation so that scores increase more at the lower end of the distribution (i.e. increasing application availability from 1,000 to 2,000 applications results in a larger score improvement that increasing from 1,001,000 to 1,002,000 applications).

Accessibility of top 400 ranked applications

This indicator focuses on the most popular applications available in a country and measures the proportion of the population that can use them, whether in their first or second languages. For each application, we estimate the proportion of the population that are able to use it based on the languages that it features. If an application is available in English, French and Hindi and 80% of a country's population speaks one of these languages (either as a first or second language), we assume that the app is accessible to 80% of the population. We then take the average of the top 400 ranked applications in each store and use the most popular store for inclusion in the Index (for example, Android is far more widely used in many markets than iOS, so for these countries, we only consider the accessibility of mobile applications on Google Play).

3. Data treatment

Having collected data and carried out the necessary calculations for the above indicators, we check to ensure that each country has data on at least 75% of indicators overall and at least half the indicators within each enabler. This ensures that a significant proportion of data for each country is based on actual data rather than estimates. We also ensure that each indicator has data for at least 75% of countries.

The next step is to then 'treat' the data, by dealing with outliers and imputing missing data. If data is skewed by certain outliers, this could impact the overall index scores (for example, a country with exceptionally low download speeds compared to all other countries will score very low but will also cause all other countries to score relatively highly with little variation). In order to identify outliers, indicators are assessed to see if they have an absolute skewness above 2 and kurtosis above 3.5[18]. Where these thresholds are met, one of two treatment approaches is adopted:

- **Winsorisation**: outlier variables are trimmed to the nearest value until the indicator is brought within the specified ranges for skewness and kurtosis. For example, if a country has an outlier value of 1,000 and the next highest value is 90, the former is trimmed to 90. If this gives acceptable skewness and kurtosis scores, the process stops there. If not, the two values are trimmed to the next highest value (which might be 80 in the above example). This process is continued until the indicator falls within the specified skewness and kurtosis ranges. In order to ensure that a large number of observations are not adjusted, a maximum of six observations are trimmed. If this is still insufficient to reduce skewness and kurtosis, the second approach is implemented.

- **Transformation**: as the majority of the indicators with high skewness and kurtosis are skewed to the right, a logarithmic transformation is used to bring the indicator within the specified ranges.

There are a few indicators where a logarithmic transformation is applied even though Winsorisation would suffice. This is because a logarithmic transformation has a conceptual benefit in that it results in improvements in the lower end of the indicator distribution being more 'beneficial' to a country than improvements at the high end of the distribution. An example of this is in relation to mobile application development. Increasing the number of applications per person from 10 to 20 is likely to have a bigger impact on providing relevant content than increasing the number from

[18] These thresholds are generally used in identifying outliers for composite indices.

100,010 to 100,020 and so – from the perspective of the Index – should be rewarded with a higher increase. Logarithmic transformation achieves this.

A logarithmic transformation has been applied to the following indicators, for either data treatment or conceptual reasons:

- International bandwidth per user;
- Number of secure servers;
- Cost of entry usage basket;
- Cost of medium usage basket;
- Cost of high usage basket;
- Cost of entry-level device[19];
- TLDs per capita;
- Mobile application development; and
- Number of applications in national language(s).

The next step in the data treatment process requires the imputation of missing data. Where data is missing, historic information is used before implementing a modelling-based approach. For data that is generally updated annually, we use the previous year's value where the latter is available and the current year is missing. This is used for indicators such as number of servers and international bandwidth per user. This process is likely to result in a more accurate estimate for a specific country than using a modelled or imputed value based on data for other countries. However, if there is no data for the current or previous year, then historic values are not used because indicators such as number of servers and international bandwidth per user can vary significantly over two years and so using data that is older than one year will be subject to greater inaccuracy.

[19] This indicator was not log-transformed in previous years as it generally fell within the specified skewness and kurtosis thresholds. However, in the last two years this has not been the case so we have applied the transformation to be consistent with the treatment of mobile data costs. While the results are consistent over time, the normalised scores for the indicator, dimension and the Affordability enabler will be different from what was published in previous years.

For some of the indicators, data is only updated every few years (or sometimes longer) – particularly if the indicator is not expected to significantly vary year-to-year and/or if collecting the data is particularly complex. This applies to the following indicators in the MCI:

- Access to electricity;
- Income inequality; and
- Education indicators and their gender ratio counterparts.

For these variables, if data exists in the period since 2014 then the most recent value is used. Otherwise, it is imputed using the methods described below.

The remaining missing data is imputed with a multivariate normal (MVN) data augmentation approach that uses multiple imputation. The MVN method generates imputed values assuming an underlying joint multivariate normal model.[20] In order to account for variation caused by missing data, the model is run 20 times – the average of these 20 imputations is then used to impute the missing value. In order to ensure that the Index rankings are robust to the imputation method, missing values were also imputed using a multiple imputation method based on predictive mean matching (PMM)[21]. We found that no countries moved more than 10 places in the 2018 rankings when using the PMM approach. This shows that the Index is not particularly sensitive to the imputation methodology used for missing data. Previous sensitivity analyses of the MCI also showed that the rankings were not sensitive to other types of imputation, for example hot-deck imputation[22].

[20] Even if the normality assumption is violated the MVN approach has still been shown to lead to reliable estimates given a sufficient sample size.
[21] This generates estimates of missing values using a regression model; the independent explanatory variables are selected if they have a high correlation with the variable being imputed. However, for a number of the indicators, imputing a value by regression produces results that are not valid – for example, negative download speeds, coverage figures greater than 100% and negative prices. A predictive mean matching (PMM) approach is therefore applied. This generates an estimated value using the regression for a country that is missing data and then matches it with the country with the closest regression output. The actual value of that country is then taken. As with the MVN approach, in order to account for variation caused by missing data, the regression is run 20 times with slightly different coefficients. The average of these 20 estimates is then used to impute the missing value.
[22] This estimates missing values by using the value of the country that is mathematically closest to it. It is implemented by identifying indicators with high correlation with the indicator with missing data. These are then used to calculate the Mahalanobis distance to all other countries. The country with the smallest distance is identified as the nearest neighbour and data is imputed using that country.

4. Normalisation

Normalisation is required in an index to adjust for different units of measurement and different ranges of variation across the indicators. For the Mobile Connectivity Index the minimum-maximum method is used, which transforms all indicators so that they lie within a range between 0 and 100 using the following formula:

$$I_{q,c} = \frac{x_{q,c} - min_c(x_q)}{max_c(x_q) - min_c(x_q)}$$

Where 'I' is the normalised min-max value, 'x' represents the actual value and the subscripts 'q' and 'c' represent the indicator and country respectively.

This method has been chosen over alternatives such as rankings and categorical scales because it retains interval-level information. For example, in the case of ranking 3G coverage, Country A might have 100% coverage, Country B might have 99% coverage, and Country C might have 90% coverage. These countries would therefore be ranked in order as 1, 2 and 3 respectively (or they may all be categorised as having the highest score on an ordinal scale). However, this ranking does not take into account the differences between countries – specifically the fact that B is much closer to A than it is to C. Furthermore, as the MCI is updated over time, a ranking- based approach may not track a country's progress as well as minimum-maximum or standardisation because a country might improve its coverage without increasing its rank.

For most indicators, the minimum and maximum used for normalisation are based on the actual minimum and maximum for that indicator, although in some cases they have been amended. For example the gender indicators, which represent female/male ratios, have a maximum threshold of one as this represents gender equality. Any country with a value greater than this is therefore not rewarded with a higher score.

To allow for comparisons of index scores over time, the minimum and maximum for each indicator are fixed. Some of the indicator maxima have therefore been adjusted where there are likely to be increases during the next few years in order to give all countries room to improve. These adjustments are based on an analysis of historic data and statistical analysis (for example, ensuring that the maxima do not significantly exceed a threshold of being two standard deviations above the mean).

As part of the normalisation process, all indicators are transformed such that they have the same orientation – i.e. a higher score always represents a ‘better’ score. This is necessary for indicators that are negatively correlated with mobile internet penetration – for example, mobile tariffs, income inequality and latency.

Previous sensitivity analyses of the MCI also showed that the rankings were not sensitive to other types normalisation, for example z-scores.[23]

[23] This transforms all indicators such that they have a mean of 0 and a standard deviation of 1.

5. Weighting

In order to aggregate indicators into dimension scores (and dimension to enabler scores and enabler scores to an overall index score), it is necessary to assign a weight to each component of the index. To construct the weights at the dimension, enabler and overall index level, a number of considerations have been taken into account, including:

- Statistical relationship between indicators and dimensions concerning mobile internet penetration;
- Analysis of consumer survey responses regarding perceived barriers to mobile internet access;
- Principal component analysis to identify weights that correct for the overlapping information implied by grouping indicators that are correlated (rather than representing a measure of importance);
- Research carried out by the GSMA and other organisations on digital inclusion and barriers to mobile connectivity; and
- Qualitative research.

Based on the above, the following weights have been used for the dimensions (**Table 3**) and enablers (**Table 4**).

Table 3: Indicator weights for dimensions

Dimension	Indicator	Weight
Network Coverage	2G network coverage	20%
	3G network coverage	40%
	4G network coverage	40%
Network performance	Mobile download speeds	33.33%
	Mobile upload speeds	33.33%
	Latencies	33.33%
Other enabling infrastructure	International bandwidth per user	30%
	Number of secure servers	30%
	Access to electricity	30%
	Number of Internet exchange points (IXPs)	10%
Spectrum	Digital dividend spectrum	45%
	Other spectrum below 1GHz	20%

	Spectrum in bands 1-3GHz	35%
Mobile tariffs	Cost of 100MB	33.33%
	Cost of 500MB	33.33%
	Cost of 1GB	33.33%
Handset price	Cost of entry-level internet-enabled handset	100%
Inequality	Inequality in income, Atkinson measure	100%
Taxation	Cost of taxation	50%
	Cost of mobile-specific taxation	50%
Basic skills	Adult literacy rate	25%
	School life expectancy	25%
	Mean years of schooling	25%
	Tertiary enrolment rate	25%
Gender equality (f)	Gender years of schooling ratio	16.67%
	Gender account access ratio	16.67%
	Gender GNI per capita ratio	16.67%
	Women, Business and the Law Index	16.67%
	Gender gap ratio for social media use	16.67%
	Gender gap ratio for mobile phone ownership	16.67%
Mobile ownership	Penetration of mobile users	100%
Local relevance	TLDs per capita	20%
	E-Government services	20%
	Mobile social media penetration	30%
	Mobile application development	30%
Availability	Number of apps in national language	50%
	Accessibility of top ranked apps	50%
Online security	Global Cybersecurity Index	100%

Table 4: Dimension weights for enablers

Dimension	Indicator	Weight
Infrastructure	Network coverage	30%
	Network performance	30%
	Other enabling infrastructure	20%
	Spectrum	20%
Affordability	Mobile tariffs	30%
	Handset price	30%
	Taxation	20%
	Inequality	20%
Consumer Readiness	Basic Skills	40%
	Gender Equality	40%
	Mobile ownership	20%
Content and Services	Local relevance	40%
	Availability	40%
	Online security	20%

In terms of weighting the enablers for the Index, equal weights are assigned – i.e. each enabler is given a weight of 25%. **Table 5** shows the Pearson and Spearman ranking correlation coefficients between the enablers and final index score against mobile internet penetration, demonstrating a high correlation across all enablers.

Table 5: Correlation coefficients with mobile internet penetration

	Pearson correlation	Spearman correlation
Infrastructure	0.88	0.89
Affordability	0.82	0.83
Consumer Readiness	0.86	0.88
Content and Services	0.84	0.86
Index	0.91	0.92

Analysis was conducted to assess the impact of adjusting these weights on the correlation between the overall index score and mobile internet penetration rates, including analysis that set weights to optimise both correlation coefficients. Such changes make very small improvements to the final index-penetration correlation (less than 0.01). Equal weights are therefore appropriate.

6. Aggregation

Once weights have been assigned to the indicators, dimensions and enablers, they need to be aggregated to produce the relevant composite scores. Two methods of aggregation were considered: arithmetic and geometric. The key consideration when choosing between these is the extent to which indicators, dimensions and enablers are substitutable, with arithmetic aggregation implying perfect substitutability and geometric implying partial substitutability.

When considering the indicators in the Mobile Connectivity Index, there is often a greater degree of substitutability compared to the dimensions and enablers. For example, within the Network Coverage dimension low 3G network coverage in a country can be compensated by high 4G network coverage in the same country. In the Mobile Tariffs dimension, an expensive price for the medium basket could be compensated by a cheap entry basket price. At the enabler level, such substitutability is unlikely to be perfect – a country with a high infrastructure score is unlikely to achieve high mobile internet penetration if mobile products or services are completely unaffordable or if there is no relevant internet content. The dimensions sit somewhere in-between – there is likely to be more substitutability than the enabler level (for example, high handset price might be compensated by a low tariff price) but less than at the indicator level (for example poor mobile coverage is unlikely to be compensated with high network quality). With this in mind, we have adopted the following aggregation rules:

- Dimension aggregation – arithmetic
- Enabler aggregation – arithmetic
- Index aggregation – geometric

7. Final Remarks

Mobile internet connectivity is driving economic, societal and individual development around the world. In many locations, mobile technology provides the only opportunity for many to get online – and, therefore, the only way for millions of people to benefit from the transformational potential that the internet can deliver.

Through measuring and tracking the enablers of mobile internet connectivity, the MCI aims to support and accelerate the mobile industry's efforts to drive mobile internet connectivity and accelerate digital inclusion. With this in mind, the GSMA welcomes any feedback regarding the MCI. This includes hearing about how the MCI is being used by others in the mobile sector, and beyond. Please contact the Connected Society team (connectedsociety@gsma.com) with any questions, comments, suggestions or citations of the MCI.