GSMA Intelligence

Mobile Connectivity Index: Methodology

July 2016

gsmaintelligence.com • info@gsmaintelligence.com • @GSMAi

Introduction

About the Mobile Connectivity Index

The Mobile Connectivity Index 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 Joint Research Centre (JRC).[1] This methodology for the Index 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; the approach to aggregation; and lastly the results of a sensitivity analysis of the Index.

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 mobile internet connectivity. An input index is therefore distinct from an output index. In the context of mobile connectivity, an output index might seek to measure the intensity and diversity of mobile internet usage. It would seek to measure and understand how (or how much) people are using mobile internet services. By contrast, an input index seeks to measure and understand why people are not using mobile internet services.

Why is an index necessary?

There is no single barrier or enabler to mobile 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 complex and multi-dimensional realities.

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

- Affordability Index (Alliance for Affordable Internet)
- Networked Society City Index (Ericsson)
- Digital Economy & Society Index (European Union)
- Global Connectivity Index (Huawei)
- Broadband Development Index (IDBA)
- ICT Development Index (ITU)

[1] Handbook on constructing composite indicators: methodology and user guide, OECD and JRC, 2008

- Barriers to Internet Adoption (McKinsey)
- Networked Readiness Index (WEF)
- The Web Index (World Wide Web Foundation).

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

- It focuses specifically on mobile connectivity rather than internet connectivity in general (including fixed). Given that the digital inclusion gap in the developing world is expected to be addressed to a significant extent by mobile, it is important to understand the enablers of mobile connectivity specifically.
- 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 134 countries that account for more than 95% of the world's population.

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 a cultural environment that promotes gender equality.
- 4) Content – the availability of online content and services accessible and relevant to the local population.

Data selection

As the Mobile Connectivity Index is an input index, it is important that each indicator is an 'input' for mobile 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 used, 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. For this version of the Mobile Connectivity Index, the majority of most recent data is available is for 2014 so this is used as the reference year.

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 such as the quality of governance and corruption levels. Although soft indicators are very useful for some indices, particularly those where hard indicators are difficult to measure, they are not used in the Mobile Connectivity Index. This is to ensure that countries 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 perfectly measure a certain enabler. For example:

- Indicators such as international bandwidth per user and fixed broadband penetration are included as proxies for the quality of a country's core network. This is because 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).

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 Index.

Table 1 presents 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
Source: GSMA Intelligence

Enabler	Dimension	Indicator	Unit	Source
Infrastructure	Mobile infrastructure	2G network coverage	% Population covered	ITU
		3G network coverage	% Population covered	GSMA Intelligence
		4G network coverage	% Population covered	GSMA Intelligence
		Years since 3G network launch	Years	GSMA Intelligence
	Network performance	Mobile download speeds	Mbps	OpenSignal
		Mobile latencies	Milliseconds	OpenSignal
	Other enabling infrastructure	International bandwidth per user	Bits per second	ITU
		Number of servers	Secure servers per 1 million people	World Bank
		Access to electricity	% of population with access	World Bank
		Fixed broadband take-up	Subscriptions per 100 inhabitants	ITU
		Fixed download speeds	Mbps	Measurement Lab
		Fixed latencies	Milliseconds	Measurement Lab
	Spectrum	Spectrum <1 GHz (a)	MHz	GSMA Intelligence
		Spectrum >1 GHz	MHz	GSMA Intelligence

Affordability	Mobile tariffs	Cost of postpaid 500 MB data plan	% of GNI per capita	ITU/World Bank
		Cost of prepaid 500 MB data plan	% of GNI per capita	ITU/World Bank
		Cost of voice call bundle	% of GNI per capita	ITU/World Bank
	Handset price	Cost of entry-level handset	% of GNI per capita	GSMA/World Bank
	Income	GNI per capita	US dollars (PPP)	World Bank
	Inequality	Gini co-efficient	Index value (0=perfect equality, 100=perfect inequality)	World Bank/CIA Factbook
	Taxation	Tax as a % of TCMO	Cost of taxation as % of TCMO	GSMA
Consumer	Basic skills	Adult literacy rate	% of adult population literate	UNESCO /CIA Factbook
		School life expectancy (b)	Years	UNESCO
		Mean years of schooling (c)	Years	UN
		Tertiary enrolment rate	%	UNESCO
	Gender equality (d)	Gender literacy ratio	Female/male ratio	UNESCO
		Gender years of schooling ratio	Female/male ratio	UNESCO
		Gender account ratio	Female/male ratio	World Bank Global Findex
		Gender labour participation ratio	Female/male ratio	ILO
		Gender GNI per capita ratio	Female/male ratio	UN

Content	Local relevance	Number of generic top-level domains (gTLDs) per capita	Number of domains per person	TLDLogic and ZookNIC
		Number of country code top-level domains (ccTLDs) per capita	Number of domains per person	TLDLogic and ZookNIC
		Quality of e-government services	Index value (0=worst, 1=best)	UN
		Facebook penetration rate	% of population	Facebook
		Wikipedia edits per user	Number of edits per internet user	Wikipedia Statistics
	Availability	Accessible Wikipedia articles for the average person (e)	Number of articles available to average person	Wikipedia Statistics and Ethnologue
		Accessible website content for the average person (f)	Number of websites available to average person	W3Tech and Ethnologue
		Average accessibility of the top 100 mobile apps to the average person (g)	Average of the % of population that can use each app in the top 100 for that country	App Annie and Ethnologue

(a) This measures the amount of spectrum that has been assigned to mobile network operators in a country at bands below 1 GHz (e.g. 800 MHz and 900 MHz). The second spectrum indicator measures spectrum assigned to operators at bands above 1 GHz (e.g. 1800 MHz and 2600 MHz).
(b) 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.
(c) 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.
(d) Each of the indicators in this dimension is calculated by dividing he relevant female indicator (e.g. female literacy) by the relevant male indicator (e.g. male literacy).
(e) In order to construct this indicator we use data on the number of Wikipedia articles and combine this with data on the languages spoken in each country. The calculation works as follows: suppose there are 5 million articles in English and 1 million in Spanish and in a specific country 70% of people speak English and 30% speak Spanish. This means that 70% of people are able to read 5 million articles and 30% can read 1 million articles. On average, a person in that country can read (5m*70%) + (1m*30%) = 3.8 million articles. If everyone in a country is bilingual and speaks English and Spanish, then the average person can read 6 million articles. The calculation therefore maps the language distribution of online content (proxied by the number of Wikipedia articles) with the languages spoken in each country.
(f) This follows a similar calculation to the Wikipedia indicator but instead uses information from W3Tech on the language distribution of the top 10 million websites accessed worldwide

(g) App Annie ranks the top 100 mobile apps in a number of countries and, for each app, has information on the language it is available in. For each app, we estimate the proportion of a country's population that can use it. For example, if it is available only in English and 30% of a country's population speaks English, then the app will be available to 30% of people. If the remaining population speaks either French or German and an app is available in those languages as well, the app will be available to 100% of people. Having calculated the availability of each app to the population, we then take an overall average (i.e. the average availability of the top 100 apps to a country's population).

Data treatment

Having obtained 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 a country is not imputed. Similarly, we also ensure there is data for at least 75% of countries for each indicator.

The next step is to then treat the data, 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 2G network coverage 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[2]. 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 still isn't sufficient 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.

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

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 GNI per capita. Increasing average incomes from $1,000 to $2,000 per year is likely to have a bigger impact on mobile affordability than increasing from $100,000 to $101,000, 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 servers per million people
- cost of postpaid 500 MB data plan
- cost of prepaid 500 MB data plan
- cost of voice call bundle
- GNI per capita
- number of gTLDs per capita
- number of ccTLDs per capita.

The next step in the data treatment process requires the imputation of missing data. For the Mobile Connectivity Index, the majority of most recent data is available for 2014 so this is used as the reference year. Where data is missing, historic information is used before implementing a modelling-based approach. For data that is generally updated annually, the 2014 value is assumed to be the same as the 2013 value where the latter is available and the former is missing. This is used for indicators such as GNI per capita and mobile prices. This 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 2013 or 2014, then historic values are not used because prices and income are likely to vary significantly over two years and so using data that is older than one year will be subject to greater inaccuracy.

For some of the indicators, the data is only updated every few years (or sometimes longer) if it 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 Mobile Connectivity Index:

- Access to electricity
- Gini co-efficient
- Education indicators and their gender ratio counterparts.

For these variables, if data is missing for 2014 but exists in the period 2009–2013 then the most recent value is used. Otherwise, it is imputed using the methods described below.

The remaining missing data is imputed using a regression-based approach that uses multiple imputation. 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.[3]

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.

To ensure the Index rankings are robust to the imputation method, missing values were also imputed using a hot deck imputation approach, which estimates data by using the value of the country that is mathematically closest to it. The hot deck imputation approach is implemented by identifying indicators that are highly correlated 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] This is slightly different to the standard approach to hot deck imputation, which generally uses all variables in an enabler (or in the Index) to calculate the mathematical distance. A revised approach was used in the Mobile Connectivity Index as, in some cases, an indicator can be highly correlated with an indicator outside its enabler. This relationship was therefore exploited in an attempt to improve the accuracy of the imputations.

If the hot deck imputation method is used, only five countries move more than 10 places in the rankings. This shows that the Index is not particularly sensitive to the imputation methodology used for missing data.

[3] For example, if the model for 2G coverage generates a predicted value for Country A of 102, the PMM method looks at the predicted value for other countries. It finds the country with the closest predicted value, say 101 for Country B. However, Country B will have actual data for 2G coverage, which might be 99.5. The model therefore imputes data for Country A as being 99.5 (i.e. using actual data rather than the regression output).

[4] We consider Mahalanobis distance to be preferable to the Euclidean distance as it takes into account the covariance structure. However, we also calculated distances using normalised Euclidean distances and generally found the same nearest neighbour.

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 min-max method is used, which transforms all indicators so 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%, Country B might have 99% and Country C might have 90%. These would 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 doesn't take into account the differences between the two – specifically the fact that B is much closer to A than it is to C. Furthermore, as the Mobile Connectivity Index will be updated over time, using a ranking approach doesn't track a country's progress as well as min-max or standardisation because a country might improve its coverage without increasing its rank.

For most indicators, the minimum and maximum used for normalisation reflect 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 1 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 will be fixed in future indices, unless there is a reasonable justification for changing it. Some of the indicator maxima have therefore been adjusted where there are likely to be increases during the next four to five years. These adjustments are based on an analysis of historic data (if there has been a general historic trend, we extrapolate this going forward to set a maximum) and statistical analysis (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 also 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, the Gini coefficient and latency.

To ensure the Index is robust to the normalisation methodology, Index scores were also calculated by normalising indicators using 'z-scores'. This transforms all indicators such that they have a mean of 0 and a standard deviation of 1. The Index rankings are robust to the normalisation method, with only one country moving more than 10 places if z-scores are used.

Weightings

To construct the weights at the dimension, enabler and overall index level, a number of considerations have been taken into account, including the following:

- Statistical relationship between indicators and dimensions with mobile internet penetration – this includes both correlation and regression analysis.
- Analysis of consumer survey responses regarding perceived barriers to mobile internet access.
- Principal component analysis – this identifies 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.
- Qualitative evidence and expert opinion within the GSMA.

Based on this, the following weights have been used for the dimensions (Table 2) and enablers (Table 3).

Table 2: Indicator weights for dimensions
Source: GSMA Intelligence

Dimension	Indicator	Indicator weights
Mobile infrastructure	2G network coverage	20%
	3G network coverage	30%
	4G network coverage	25%
	Years since 3G network launch	25%
Network performance	Mobile download speeds	50%
	Mobile latencies	50%
Other enabling infrastructure	International bandwidth per user	20%
	Number of servers	20%
	Access to electricity	25%
	Fixed broadband take-up	15%
	Fixed download speeds	10%
	Fixed latencies	10%
Spectrum	Spectrum <1GHz	65%
	Spectrum >1GHz	35%
Mobile tariffs	Cost of postpaid 500 MB data plan	40%
	Cost of prepaid 500 MB data plan	40%
	Cost of voice call bundle	20%
Handset price	Cost of handset	100%
Income	GNI per capita	100%
Inequality	Gini co-efficient	100%
Taxation	Tax as a % of TCMO	100%
Basic skills	Adult literacy rate	25%
	School life expectancy	25%
	Mean years of schooling	25%
	Tertiary enrolment rate	25%
Gender equality	Gender literacy ratio	30%
	Gender mean years of schooling ratio	30%
	Gender account ratio	20%
	Gender labour participation ratio	10%
	Gender GNI per capita ratio	10%
Local relevance	Number of gTLDs per capita	25%
	Number of ccTLDs per capita	25%
	E-government services	25%
	Facebook penetration rate	15%
	Wikipedia edits per user	10%
Availability	Accessible Wikipedia articles	10%
	Accessible website content	10%
	Average accessibility of the top 100 mobile apps	80%

Table 3: Dimension weights for enablers
Source: GSMA Intelligence

Enabler	Dimension	Dimension weight
Infrastructure	Mobile infrastructure	30%
	Network performance	30%
	Other enabling infrastructure	20%
	Spectrum	20%
Affordability	Mobile tariffs	20%
	Handset price	20%
	Income	20%
	Inequality	20%
	Taxation	20%
Consumer	Basic skills	50%
	Gender equality	50%
Content	Local relevance	50%
	Availability	50%

In terms of weighting the enablers for the Index, equal weights are assigned – i.e. each enabler is given a weight of 25%. Table 4 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 4 Correlation coefficients with mobile internet penetration
Source: GSMA Intelligence

Enabler/index	Pearson correlation	Spearman correlation
Infrastructure	0.86	0.87
Affordability	0.81	0.82
Consumer	0.75	0.77
Content	0.83	0.84
Final index score	0.88	0.90

An analysis was carried out 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.

Aggregation

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.

At the lower levels of the Mobile Connectivity Index, there is often a greater degree of substitutability than at the higher levels. For example, within the Mobile Infrastructure dimension low 3G network coverage can be compensated by high 4G network coverage. In the Mobile Tariffs dimension, a high postpaid price plan can be compensated by a low prepaid price plan. At the index 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 is completely unaffordable or if there is no relevant content. The enabler groups sit somewhere in-between – there is likely to be more substitutability than the index level (e.g. high handset price might be compensated by a low tariff price) but less than at the dimension level (e.g. poor mobile coverage is unlikely to be compensated with high network performance). With this in mind, we have adopted the following aggregation rules:

- dimension aggregation – arithmetic
- enabler aggregation – arithmetic
- index aggregation – geometric.

If geometric aggregation is used at the enabler level, four countries move more than 10 places in the rankings. If arithmetic aggregation is used at all levels, no countries move more than 10 places. This shows that the Index rankings are robust to the method of aggregation.

Sensitivity and uncertainty analysis

In Table 5, the results of the following sensitivities are presented:

- imputation method: hot deck instead of multiple imputation
- normalisation: z-scores instead of min-max
- aggregation: one scenario where everything is arithmetic and another where enabler aggregation is geometric.

The impact of combining all three adjustments is also presented.[5] The impact is assessed by looking at the number of countries that shift more than 5, 10 and 15 positions in the rankings.

[5] In this scenario, the aggregation assumption is all arithmetic as geometric cannot be used with z-scores

Table 5: Sensitivity analysis of index
Source: GSMA Intelligence

Adjustment	Number of countries moving more than		
	5 places	10 places	15 places
1. Hot deck imputation	10	5	4
2. Z-scores	12	1	1
3. All arithmetic aggregation	5	0	0
4. Geometric aggregation for enablers	21	4	4
1, 2 and 3 combined	19	5	5

The analysis shows that the Index is robust to a change in assumptions, with few countries moving more than 10 places in the rankings depending on the adjustment. The impact of having geometric aggregation at the enabler level impacts some countries quite significantly in the lower half of the rankings as it punishes poor performance at the dimension level more heavily. However, the overall proportion of countries materially affected remains fairly small.

An additional sensitivity regards the weights that are applied. As discussed above, adjusting weights at the enabler level does not make a material difference to the Index correlation with mobile internet penetration. Further analysis shows that adjusting the enabler weights by up to 5 percentage points does not shift any country rankings more than 10 places. The Index is therefore also robust to the weights assumed at the enabler level.

About GSMA Intelligence

GSMA Intelligence is the definitive source of mobile operator data, analysis and forecasts, delivering the most accurate and complete set of industry metrics available.

Relied on by a customer base of over 800 of the world's leading mobile operators, device vendors, equipment manufacturers and financial and consultancy firms, the data set is the most scrutinised in the industry.

With over 30 million individual data points (updated daily), the service provides coverage of the performance of all 1,400+ operators and 1,200+ MVNOs across 4,500+ networks, 77 groups and 238 countries worldwide.

gsmaintelligence.com • info@gsmaintelligence.com • @GSMAi

GSMA Intelligence, GSMA, The Walbrook Building, 25 Walbrook, London EC4N 8AF