

CAPITA

**Paper on the Statistical Information Used in
Contact-Centres**

Tabled by the Capita Group

Produced by Eamon Goodfellow

Issued 30th October 2003

Table of Contents

1	Phone Performance Standards	4
1.1	% Calls Abandoned	4
1.2	Average Speed to Answer (ASA)	8
1.3	% Of Calls Answered in a Target Time.....	9
2	Agent Statistics	12
3	Three Key Measures	13
3.1	Occupancy	13
3.2	Utilization	13
3.3	Productivity	14
4	Table of Further Contact-Centre Calculations	16
5	Outbound.....	18

Background

When I was first approached to do this by the CCA the premise was quite simple – create a standard list of measures to be included in a pack of best practice principles to be published by the CCA. Little did I know at the time that the complexity and lack of uniformity across the industry would result in much debate in my mind as to the best way forward on this. Essentially what I have done is to try and simplify the process of measurement to a point where it will at least be familiar to all companies that have a contact-centre element. I have also proposed a number of calculations which I use for measuring performance and assisting with resource forecasting. I have no doubt that some of you will not agree with the calculations I have defined, for others it will provide thought and debate and for still others it will provide a starting point for performance management, it does however provide frame work around which all contact centres can be measured and similarly compared.

First, some background. I started in contact centres as an agent in BBC Information in Belfast (a Capita run contract) in 1998. I started helping out with the daily statistical reporting until my line manager announced one morning that he could not go on a course for Resource Planning and sent me instead. Within a couple of weeks I was responsible for the resource management of the contract and completing the service level reporting for our client the BBC, following this I moved on to be Resource Manager for the Criminal Records Bureau. In 2001 I left Capita to take up the position of MIS / Resource manager with Churchill Insurance before returning to Capita's sales team, working on contact centre bids and looking at our current contact centre provision.

Throughout I have been attempting to “square the circle of reporting” that exists in all contact-centre set-ups i.e. matching the people required against the people available and then reporting on the performance afterwards.

Workforce Management Tools (WFM) and bespoke Erlang spreadsheets are now essential to the resource planning of any contact centre. ACD's route the calls in an ever more complex and efficient manner to agents who are monitored every minute of the day that they are logged in. The difficulty I have found is in tying all of these elements together so that the forecast staffing requirement results in an actual staffing number in volume and profile that matches the incoming volume of work, then reporting on their activity.

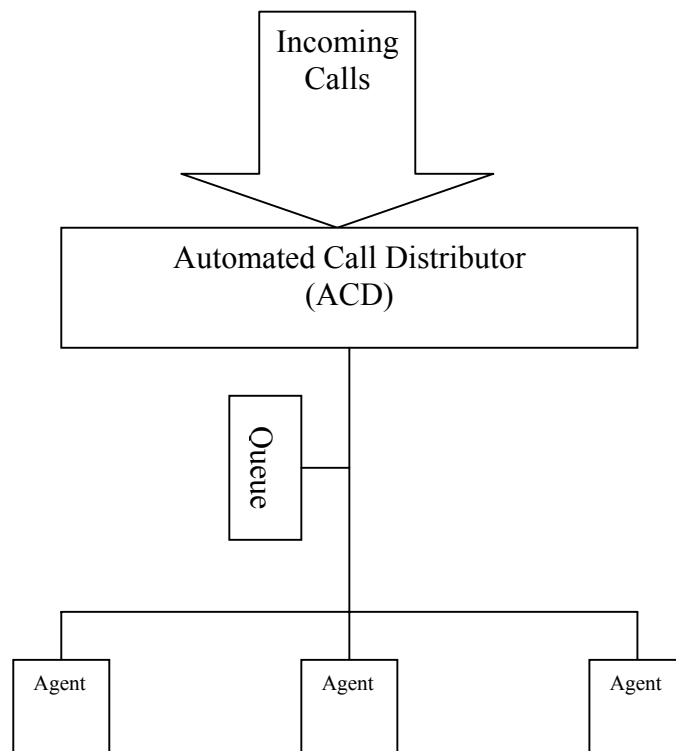
When I first started resource planning I thought it was easy as I had vastly more people than needed to answer the calls, then I started factoring in Annual leave, then sick leave, then meetings, breaks, 1-2-1's etc to the point where there were now parts of the week when we didn't have enough people, so I have to re-juggle the schedules and plan overtime. It was only afterwards when I started analysing performance that I saw that in some periods when we had, on paper, the right number of people and the right average transaction time we still didn't hit our service targets. It was at this point that I started to formulate ideas on how to view performance in contact centres in order to get the resource planning right.

1 Phone Performance Standards

Most, if not all contact centres have a phone performance standard stipulated internally or as part of a service level agreement with their client. These standards are intended to drive performance to a certain level of customer expectation whilst also forming an integral part of the metrics required to calculate staffing levels. In most centres the standards applied are-

- % Calls Abandoned
- % Calls answered in x seconds
- Average speed of answer (ASA)

The below diagram shows a very simple telephone set-up for a call-centre



In this set-up incoming calls (commonly referred to as “OFFERED”) are passed to the ACD which then searches for an available agent (agent is a common term used to describe staff logged on to the ACD system to handle calls) to connect the call to. If it cannot find an agent, (e.g. all the agents are busy handling other calls) it will place the call in a queue. Once there the caller will either wait until an agent becomes available or hang-up (ABANDON). When connected to an agent the call is referred to as ANSWERED and the ACD notes how long it took for the call to be ANSWERED by an agent from the time it was OFFERED to the ACD.

From this set-up we can derive the three most common service standard calculations.

1.1 % Calls Abandoned

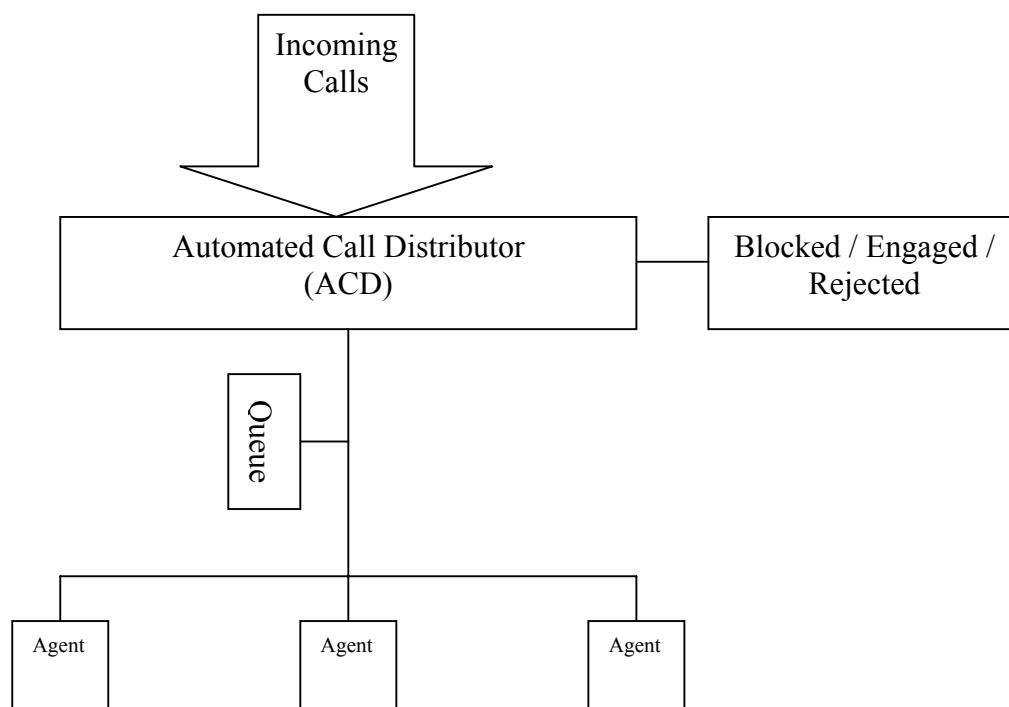
100 calls are OFFERED to the ACD over the period of a day, of which 20 are sent to the queue, as an agent is not immediately available. Of the 20 calls that were sent to the queue 15 callers waited until they were connected to an agent and 5 hung-up (ABANDONED).

% Calls Abandoned is - $\frac{\text{Number of calls ABANDONED in queue}}{\text{Number of calls OFFERED}}$

$$5 / 100 = 5\%$$

It is at this point that I have to introduce the first of a number of complexity elements which will become a feature of this paper – ACD Capacity

ACD Capacity describes the number of simultaneous calls that an ACD can handle at any one time. Consider the set-up I have used to describe abandoned calls. If the ACD has a capacity of 20, this means that it can handle 20 simultaneous calls, be they in the queue or being answered by an agent. Effectively an incoming call uses 1 part of the ACD's capacity from the time it is offered to the ACD until the call is concluded (i.e. caller or agent hangs up). If the ACD is already handling 20 calls the next call offered will receive an engaged tone, whilst still recorded as offered*. The set-up will then look like this



This means that the calculation for % Calls Abandoned now has an added layer of complexity. Any calls receiving the engaged tone (also referred to as BLOCKED or REJECTED) from the ACD should be excluded from the calculation for % Calls Abandoned. This is because we are trying to measure the % of calls which abandon from all those who have the potential to reach that point in the set-up, this can then indicate the effectiveness of queuing messages. The calculation for % Calls Abandoned becomes

$\frac{\text{Number of calls ABANDONED in queue}}{\text{Number of calls OFFERED} - \text{calls BLOCKED}}$

It is however important to measure any blocking of calls at the ACD level. This can be simply expressed as

* Those amongst you screaming about network delivery, overflows, reserve capacity plans etc will have to be patient at this point, we'll get to those more complex arrangements later.

% Calls Blocked - $\frac{\text{Number of calls BLOCKED at ACD level}}{\text{Number of calls OFFERED}}$

Short Calls Abandoning

A common problem in most call-centres is their abandonment rate being affected by short abandoned calls e.g. callers dialling the wrong number, or being interrupted whilst waiting to be answered, or listening to a queuing message of a phone ringing. Most contracts now reflect that these calls can have an adverse effect on reporting which is not immediately in the control of the call-centre. This is done by removing calls which abandon in a certain period of time (usually < 10 seconds) from any abandonment rate calculation. Consider the following five calls, these five calls all abandoned in a period where the number of calls offered was 100 and no calls were blocked.

Call Number	Outcome	Time in Queue (secs)
1	Abandoned	2
2	Abandoned	3
3	Abandoned	12
4	Abandoned	15
5	Abandoned	20

Without taking short calls into account the abandonment rate would be

$$\frac{\text{Number of calls ABANDONED in queue}}{\text{Number of calls OFFERED} - \text{calls BLOCKED}}$$

$$= \frac{5}{100 - 0}$$

$$= 5\%$$

Additionally this provides the inverse for a calculation looking at the % calls Answered

$$\% \text{ Calls Answered} = \frac{\text{Number of calls ANSWERED by an Agent}}{\text{Number of calls OFFERED} - \text{calls BLOCKED}}$$

$$= \frac{95}{(100 - 0)}$$

$$= 95\%$$

Looking at the call details we can see that 2 calls abandoned within 10 seconds. To take them into account the % **Calls Abandoned** becomes

$$\frac{\text{Number of calls ABANDONED in queue} - \text{Calls Abandoned in 10 seconds}}{\text{Number of calls OFFERED} - \text{minus calls BLOCKED} - \text{Calls Abandoned in 10 seconds}}$$

The common mistake made in this calculation is that calls Abandoned in 10 secs or less are removed from the calls Abandoned figure but not from the calls Offered figure. It worth remembering that excluding calls that abandon by a certain time means excluding them

from the whole of the calculation, not just the part that will make the service level look lower.

Once a call has reached the stage where it can be either answered straight away by an agent or placed in a queue to be answered is also where we should measure **Average Speed to Answer (ASA)** and **% of Calls Answered in Target**

1.2 Average Speed to Answer (ASA)

Simply put this is the average amount of time it takes for a call to be answered once the ACD has either presented the call to an agent or placed it in a queue. Say Five calls are OFFERED to the ACD. The first two calls are answered immediately by the agents. Call number 3 has to wait 10 seconds in the queue to be answered, call 4 – 15 seconds and call 5 – waits for 20 seconds in the queue before the caller decides to hang up (ABANDONS).

Call Number	Outcome	Time in Queue (secs)
1	Answered	0
2	Answered	0
3	Answered	10
4	Answered	18
5	Abandoned	20

As we are measuring the AVERAGE SPEED TO ANSWER, it is only answered calls that we should take into account, therefore the total wait time for the four calls that were answered is 28 seconds (0 + 0 + 10 + 18), the total number of calls answered is 4, so

$$\begin{aligned}
 \text{Average Speed to Answer} &= \frac{\text{Total wait time for all calls that were answered}}{\text{Total number of calls answered}} \\
 &= 28 / 4 \\
 &= 7 \text{ seconds}
 \end{aligned}$$

This calculation is also sometimes known as the Average Delay

Along with the Average speed to answer it is important to quote the average time to abandon. Consider these five calls

Call Number	Outcome	Time in Queue (secs)
1	Answered	0
2	Answered	0
3	Abandoned	20
4	Abandoned	30
5	Abandoned	40

Calls 3, 4 and 5 are those which abandon in 20, 30 and 40 seconds respectively. Therefore

$$\text{Average Time to Abandon} = \frac{\text{Total wait time for all calls that abandoned}}{\text{Total number of abandoned calls}}$$

1.3 % Of Calls Answered in a Target Time

This is the most popular measure used to judge contact-centre performance. It is also referred to as the Call-handling Standard, Service Level or Grade of Service. This measure looks at the same figures as the ASA calculation but instead measures the percentage of calls answered in a particular target time. This measure is used alongside the abandonment rate in most contact-centres to measure performance.

This calculation does though cause a certain amount of debate in contracts. The debate surrounds whether abandoned calls should be used in this calculation as well, I'll explain – say your centre is aiming to answer 90% of calls in 20 seconds, what exactly does that mean. 90% of calls offered, answered, dialled???? Answered by whom or what, an agent, someone taking a message, a call-back system, an IVR? My experience in contracts has always been to ensure clarity and measure accurately the customer service provision. The key in this is to ensure that you are measuring something is accurate and relevant, say you are trying to answer 90% of calls in 20 seconds, based on the following information.

Calls Offered	102
Calls Blocked	2
Calls Answered by an agent	95
Calls Abandoned	5
Calls Abandoned in 10 seconds	1
Calls Answered in 20 seconds	90

Using the calculations described previously we know that

$$\% \text{ Calls Blocked} = 2 / 102 = 1.96\%$$

$$\% \text{ Calls Answered} = 95 / (102 - 2) = 95\%$$

$$\% \text{ Calls Abandoned} = 5 / (102 - 2) = 5.00\%$$

$$\% \text{ Calls Abandoned (taking short calls into account)} = (5 - 1) / (102 - 2) = 4.00\%$$

I looking at the % of calls answered in 20 seconds though we can create 4 different formulas

Divide Calls Answered in 20 seconds by Calls Offered

$$= 90 / 102$$

$$= 88.24\%$$

Divide Calls Answered in 20 seconds by Calls Offered minus Calls Blocked

$$= 90 / (102 - 2)$$

$$= 90.00\%$$

Divide Calls Answered in 20 seconds by Calls Offered minus Calls Blocked minus Calls Abandoned in 10 seconds

$$= 90 / (102 - 2 - 1)$$

$$= 90.91\%$$

Divide Calls Answered in 20 seconds by Calls Answered by an agent

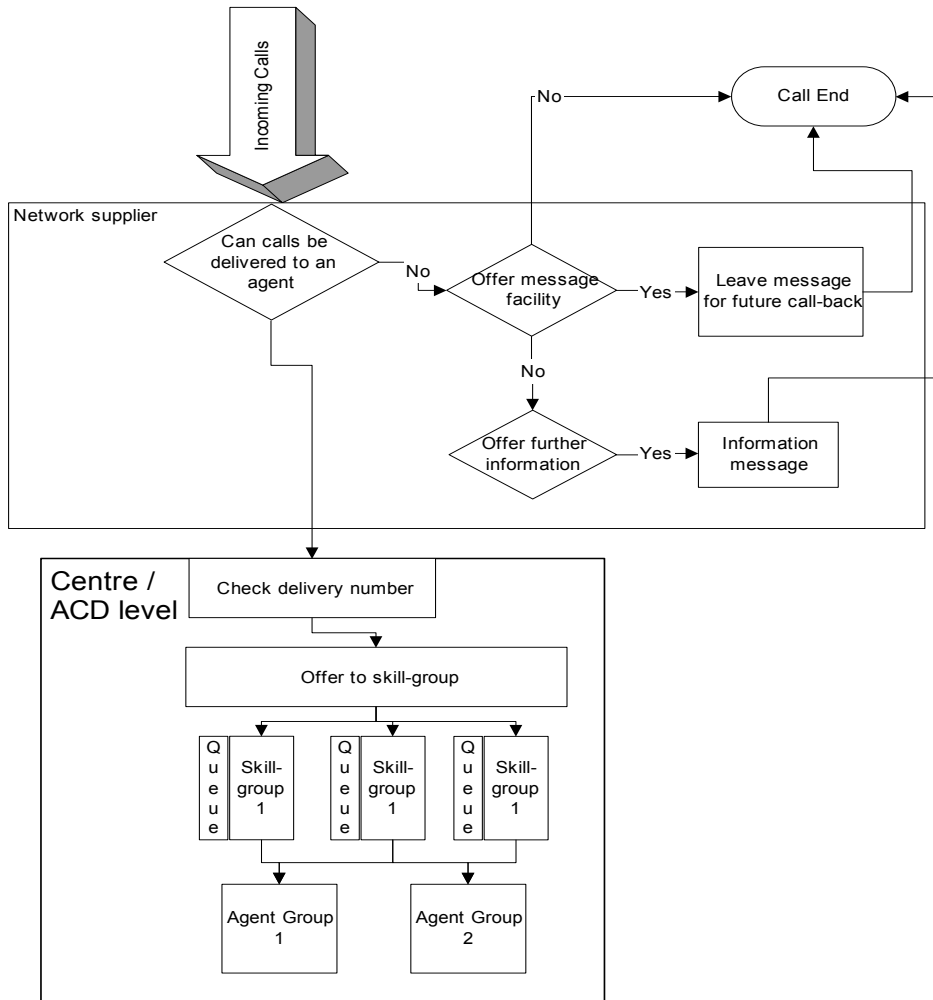
$$= 90 / 95$$

= 94.74%

The problem here is that if the agreement between a service provider and client isn't clearly set out then there can be a wide range of results, some where service level is achieved and some where it isn't. The key is obviously to agree beforehand the calculation to be used, however of the calculations shown above the fourth one is the clearest as it measures the speed of response eliminating the effects of queuing messages, line capacity and short calls. You should though use all of the calculations described in this section to get an accurate picture of the customer service provision.

More Complex set-ups

Most contact centres have a more complex set-up than the ones I have described above. Sadly this can mean sometimes that reported results are based on the terminologies used by the various software providers rather than what is actually agreed between service provider and client.



The calculations shown relate to those which need to be reported on to the level of detail required by the centre that could be one overall figure or a series of individual figures relating to individual services. The diagram above shows a more complex system but the theory remains the same. Calls Offered can be collected at the incoming call stage at the first point of Network usage, from this calls can be stripped out to show which have been dealt with at the network stage, this then gives a net calls offered figure for those calls to the centre itself for the purposes of calculating abandonment rates and speed of answer etc.

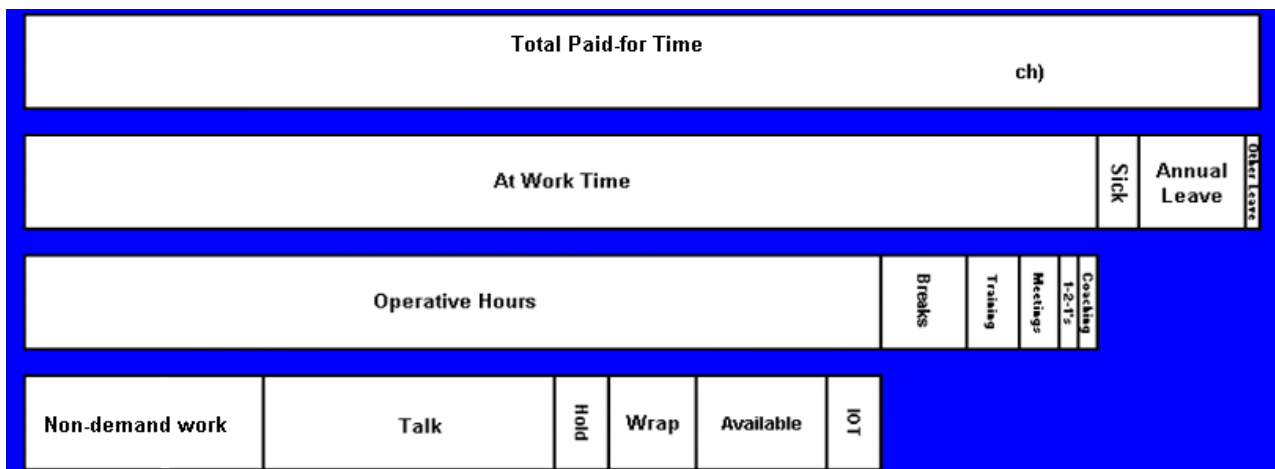
The design of the system of call handling is important as it defines the entry and exit points of each call, from this we need to find out the volume that hits each point and apply calculations to those points in order to judge the customer service provision.

2 Agent Statistics

So far we have looked at statistics which show the customers experience in dealing with a contact centre. We now need to look at how staff in the centre perform against the incoming volume.

All staff in contact centres have various states in which they exist in their working lives, they are either in work or on holiday, taking a call or on a break etc. Indeed this not only applies to contact centres but to all workplaces. The advantage we have in contact centres is that due to the interaction between the agent and either ACD, workforce management software or CRM applications we can easily extract this information, the question is what we do with it.

The diagram below shows the various states a contact-centre agent will exist in



Definitions

Non-demand work is work such as e-mails, letters, SMS, web contacts etc.

IOT is Indirect Operational Time, time when an agent is scheduled to be handling incoming work but is doing something else e.g. going to the toilet, getting a drink etc

It is this collection of data along with the number of work items dealt with which is the core information required. It is also important to take into account the skills base of the agents involved i.e. agents may be split in to discreet skills groups and handle only inbound call traffic or only e-mails or be multiskilled.

The table below describes indicative percentage Indicative figures for a steady-state contact-centre operation.

State	% Of Total Paid for Time	Notes
Sick	5.0%	
Annual Leave	9.1%	23 days out of 252
Other Leave	0.4%	1 day out of 252
At Work Time	85.5%	of FTE
Breaks (assumes unpaid Lunch)	6.7%	0.5 hours in an 7.5 hour day
Training	2.0%	5 days each year
Meetings / briefings / e-mails	2.7%	1 hours per week
1-2-1's	0.7%	1 hour per month
Coaching	1.3%	0.5 hours per week
Operative Hours	72.2%	of FTE
IOT / Not ready / Unavailable / FAT	3.6%	3 mins in the hr / 5% of Operative hours
Non-demand Work	3.6%	5% of Operative hours
Talk	45.5%	63% of Operative hours
Hold	2.9%	4% of Operative hours
Wrap	6.5%	9% of Operative hours
Available	10.1%	14% of Operative hours

Collecting this data allows us to formulate three important calculations for contact centres

3 Three Key Measures

3.1 Occupancy

This measure looks at both the amount of pressure the telephone agents are under and can indicate the effectiveness of resource scheduling. It is calculated by taking the total amount of time agents spend in Talk, Hold and Wrap and dividing it by the sum of Talk, Hold, Wrap and Available. Outbound Calls can also be included in this calculation if a predictive dialler is in place. Web-chat could also be included, as a demand activity, though the number of consecutive web sessions allowed per agent should be considered alongside the resulting Occupancy figure.

Occupancy is an excellent measure to use as it falls at a natural level given the average transaction time, required service level, number of agents scheduled to handle that activity and the technology set-up. This means that it is not always possible to raise the level of Occupancy, especially in those environments where you have long average transaction times and / or proscriptive Service level targets, without making fundamental changes to the type of work that the agents are doing or how they are scheduled to do that work.

3.2 Utilization

This measure looks at the time that an agent or agent body spends in work related activities as a percentage of their total paid for time. The table below splits the various states in work and non-work related activities.

Work Related	Non-work Related
Non-demand work	Other Paid Leave
Talk	Annual Leave
Hold	Sick
Wrap	Paid breaks
Training	Available
Meetings	IOT
1-2-1's	
Coaching	

All of the work related activities can be said to have a “positive” effect and contributing to the business, whilst the Non-work related activities can be said to have no “positive” effect on the business.

Utilization is therefore calculated as Work Related activities divided by the sum of all activities (Work and Non-work Related).

3.3 Productivity

Productivity is a word that covers a myriad of different measures and calculations in business. My suggestion is make this measure directly related to the work completed and the total paid time for an agent or agent body.

The calculation would therefore be the sum of Non-demand time, Talk, Hold and Wrap divided by the Total paid-for time. This calculation then gives us a direct relationship between the amount of work incoming and the number of FTE that need to be employed to complete that work.

This is especially useful for contact-centre sizing, providing your forecasting is correct, you can apply the percentage figure in an existing similar set-up to a proposed new centre or an increase / decrease in volumes.

For example, in the shown table of figures productivity comes out at 58.5%

Non-demand Work	3.6%
Talk	45.5%
Hold	2.9%
Wrap	6.5%
Total	58.5%

Say you're sizing a call centre answering 1 million calls per year at 3 millions calls per year, how many FTE would you need?

The total incoming workload, in seconds would be 1,000,000 x 180 = 180,000,000 secs
 This figure equates to 58.5% of the total number of FTE that need to be employed

180,000,000 / 58.5% = 307,692,308 secs of total paid for time required
 That equates to 5,917,160 per week (divided by 52)

If you are employing staff at an FTE standard of 35 hours per week then the number of FTE required is - $5,917,160 / (35 \times 60 \times 60) = 46.96$ FTE!

4 Table of Further Contact-Centre Calculations

The following table shows further calculations that can be used in contact-centres.

Definition
<p>Occupancy It is calculated by taking the total amount of time agents spend in Talk, Hold and Wrap and dividing it by the sum of Talk, Hold, Wrap and Available. (See above)</p>
<p>Utilization Calculated as Work Related activities divided by the sum of all activities (Work and Non-work Related). (See above)</p>
<p>Productivity The calculation would therefore be the sum of Non-demand time, Talk, Hold and Wrap divided by the Total paid-for time. (See above)</p>
<p>Average Handle Time Per Transaction (AHT). The average amount of time that it takes an agent to fully perform one transaction from when it arrives with them until when they are next available to deal with another contact. Specifically for call agents this would include talk, hold, and wrap (after-call work directly associated with the call that has just finished).</p>
<p>Availability Directly related to Occupancy in that %Occupancy + %Availability = 100%. Both Availability and Occupancy can be used to see how effective resource management, the pressure that agents are under and how attainable a given service level is, for instance it may sometimes be necessary for extra resource to be added in order to inflate availability to the point where a service level can be reached.</p>
<p>Schedule. Also referred to as a shift or tour. The plan of work for an agent or group of agents over a given period of time. This includes all work and non-work activities planned for that agent / agent group.</p>
<p>Adherence To Schedule A general term that refers to how well agents adhere to their schedules. Measure in two ways - Total Adherence and Accuracy of Adherence</p> <p>Total Adherence. Measures the amount of time that an agent spends in a particular "state" (call-time, lunch, non-phone work etc.) against the planned time according to the schedule.</p> <p>Accuracy of Adherence. Measures how closely an agent sticks to their schedule, did they go on breaks at the planned time? This can be measured as the percentage of time that agent spends out of adherence as a total of their scheduled time.</p> <p>It should be noted that in some businesses an agent's schedule can be very fluid so it is important to measure adherence against an agent's final schedule. E.g. it may have been necessary on the day to change an agent's scheduled lunch, that agent's adherence should be measured against the change to their schedule not the original plan</p>
<p>Call Load Also referred to as Work Load. Call Load is the product of (Average Handle Time) x call volume, for a given period.</p>
<p>Cost Per Call Total costs (fixed and variable) divided by total calls for a given period of time.</p>

Definition	
Full-Time Equivalent (FTE)	A term used in scheduling, forecasting and budgeting, whereby the number of scheduled hours is divided by the hours in a full working week. A working week will differ business to business depending on what hours full-timers are paid for in that business (35, 37.5, 40 etc) The hours of several part time agents may add up to one FTE.
Total Paid-for Time	The total amount of time in a period for which an agent is remunerated, this can include sickness, annual leave, paid breaks (not usually lunch) and all work activities.
% First Time Call Resolution	The total number of unique CLI's divided by the total number of calls where a CLI was presented over a given period of time.
Attrition	<p>Shown as a total and split into three sub-sections</p> <p>Forced Attrition – covers leaving reasons such as Redundancy, sacked, failed probationary, contract not renewed etc</p> <p>Unforced Attrition (Internal) – covers leaving reasons such as promotion, changed jobs all within the same company</p> <p>Unforced Attrition (External) – covers leaving reasons such as promotion, changed jobs all outside of the company</p> <p>In each case, the calculation should be the total number of people (not FTE) who left the company in a given period divided by the sum of the number of people employed by the business at the start of the period plus the number of people recruited into the business in that same period.</p> <p style="text-align: center;"> Staff Attrition = $\frac{\text{Total number of leavers in the period}}{\text{Headcount at the start of the period} + \text{Heads recruited during the period}}$ </p>
Staff Turnover	<p>Expressed as a percentage</p> <p style="text-align: center;"> Staff Turnover = $\frac{\text{Total number of Leavers in the period}}{\text{Average headcount in the period}}$ </p>

5 Outbound

What follows is a list of metrics supplied by Bibi Bajwa of Teleperformance which relate specifically to Outbound campaigns

Definition
<p>List Penetration</p> <p>The percentage of leads that have been touched by the dialler at least once</p>
<p>List Utilisation</p> <p>The percentage of leads against the total list size which have been completed i.e. disposition code, call backs which have reached the set threshold</p>
<p>Contact Penetration</p> <p>The percentage of effective contacts over the total list size</p>
<p>Wait Time</p> <p>The time from when an agent becomes available to take another call, to when a call is connected to them.</p>
<p>Wrap Time</p> <p>When an agent has finished talking to a called party or record (i.e. answer phone), but is still updating their details before becoming available to take another call.</p>
<p>Talk Time</p> <p>When an agent is talking to a called party.</p>
<p>Abandon Call</p> <p>This is a connect made by a dialler for which no agent is available, leading to abandonment of the call by the dialler. The Abandon Call Rate is usually defined as total number of calls where a connect was made but no agent available divided by the total number of calls where the dialler made a connect</p>
<p>Dialler Hour</p> <p>Made up of Talk, Wait and Wrap</p>
<p>Recycling Rules</p> <p>The rules that have been put in place governing when and how often types of disposition codes will be tried by the dialler</p>