

# Service Level Measures in the Supply Chain

A White Paper by Opex Analytics

## Purpose of Document

This document is meant to help explain different service level measures in the supply chain. Specifically, we are going focus on the measures from the supplier's viewpoint when shipping to its customers. We are also assuming that the supplier's customers are other businesses. So, this measure may be used by CPG (consumer packaged goods) companies shipping to retailers or a supplier shipping product to other manufacturers.

We want to provide clear definitions on service levels, how a company can evolve from one metric to another, and some common pitfalls in the implementation.

We are also writing this document to fill a gap in available documents on the topic. We found an excellent article by Dan Gilmore from Supply Chain Digest ([here is the link](#)-- it is worth a read) on the Perfect Order Metric. He also references Kate Vitasik who published some results on the perfect order. We will borrow from Gilmore and Vitasik's ideas, but they only cover the perfect order. It should be noted that our search was not extensive and if there are other good sources, please let us know and we'll include them in this document.

It is also important that the reason you want to measure (and improve) service level is to help your company make more money-- by keeping your current customers happy and continually ordering from you, by winning new customers with a competitive service level, and by avoiding the extra cost associated with poor service level-- like expediting, having extra people to deal with angry customers, with having to touch each order more times than needed, etc.

## Timeline of an Order

To understand how service level is measured, it is important to understand the timeline of events and what is being measured.

The following are the general steps. All companies will be different, but this should give you a good framework.

1. The customer issues a PO (purchase order) to the supplier. The PO is a request for a certain quantity of specific items (typically called a SKU) with a due date. If multiple different SKUs are ordered, they are listed on different lines of the PO and are often

called, obviously enough, line items. For example, the PO may request 100 units of SKU123 and 200 units of SKU456 with a due date of Oct 16.

2. Once the supplier receives the PO, they may add a shipment date to the PO. The shipment date is when the order must ship from the supplier's site to reach the customer on-time.
3. The supplier checks to see if the items are in stock or will be ready (either purchased or produced) in time when they are picked.
4. If items need to be made or procured, they will schedule this activity.
5. The supplier schedules when the orders will be picked at the warehouse (or factory). And, if the supplier doesn't have the full amount, will decide how much is shipped
6. The supplier schedules the truck to be ready to be loaded on a certain date
7. The supplier then makes or orders the items (if they are not already in stock), picks the item from the warehouse, loads the items onto the truck, and then sends the truck on its way to the customer.
8. The truck arrives at the customer and the customer inspects the items to ensure they received what they should have and checks the products for damage.

## The Evolution of Good Service Level Measures

Given the above timeline, you start to see what you might like to measure: did you ship the quantity that the customer requested, did you ship it on time, did it arrive on time.

Since some of these measures are easier to track and report on, good service level measures have evolved from the simple to the more complicated. Here are the three basic measures, each adding to the previous one.

1. **Unit Fill Rate** (or often called Case Fill Rate in the CPG industry). This represents the percent of the quantity ordered that shipped on time. Note that this measure counts on-time from when it shipped, not when the customer requested it. It is measured this way because the supplier may have more control over when the order ships rather than when it arrives. For example, if a supplier has two PO's on a given day one with quantity of 600 and the other with 400. If both ship on time with the full 1,000, you have a 100% fill rate. If both ship, but you only ship 990 items, your fill rate is 99%. If the one for 600 ships on time, but the 400 ships a day late, the fill rate is 60%.
2. **On Time In Full (OTIF)**. The OTIF measure is like the Unit Fill Rate except it counts the order as on time if it arrives at the customer on-time. That is, it no longer uses when it ships, but when it arrives. This is clearly a better measure, since it is the service level that your customer sees. However, it can be difficult to capture confirmation from either the trucking company or the customer that the order actually arrived on time. This is why this measure is listed 2nd in the evolution of service level
3. **Perfect Order**. The Perfect Order is OTIF, but typically tracks whether the order arrived damage free and if the invoice and labelling were all correct. (See [Dan Gilmore's article](#) for a more detailed discussion.) This one is listed after OTIF because it is again more

difficult to measure. It can be difficult to get information on product damage and correctness of the invoice from the customer.

You should think about these as the basic measures. There are plenty of nuances that you may want to consider for your specific calculations. For example, the following list are some of issues that you might want to consider.

- **Pick the right unit of measure.** In the example above for Unit Fill Rate, we assumed an order of 600 units and another of 400 units. It makes sense to add these units together if everything you make is very similar. For example, if the 600 units was for canned peas and the 400 for canned carrots, then it is easy to add them together. However, if you make industrial printers and supplier and one order was for 600 cases of printing paper (each worth \$25), and the other was for 400 high end printers (each worth \$4,000), it becomes more problematic to add the 600 to 400. Instead, you need to think about how to normalize the units. Clearly, being late on the 400 high end printers will have much larger impact on your business than cases of paper.
- **Decide how much partial credit you get.** This issue becomes important with OTIF and Perfect Order measures. For example, in OTIF, if the customer orders 100 units and you ship them 99 units that arrive on time, do you give yourself a service level of 99% (you got 99% of the items there on time) or 0% (they ordered 100 and you only shipped them 99). There is no right answer here. The latter measure obviously puts a lot more pressure on you to ship the full amount. But, it may also make service look a lot worse than what the customer perceives. For example, if the customer really cares that the order arrives on time and doesn't care that you missed a few items, then the former may be better. If the customer really wants everything or nothing, then the latter may be more appropriate.
- **Think about line items.** Line items are related to the two issues above on units of measure and partial credit. We've often seen companies measure service level based the number of line items of a PO that are filled on time and complete. That is, if you fill all the quantity of a line item, you get 100% credit for that line. If you miss the units by 1, you get 0% credit. This can lead to gaming the system by filling the line items with a small number of units, and, even worse, once you realize that you will miss a line, you decide to ship 0 of that line item rather than 95% of it. This latter example can be troubling if that line item represents a lot of revenue. Also, line item measures suffer from the same unit of measure problems listed above-- some items are worth \$20 per unit and some are worth \$4,000 per unit.
- **Decide what counts as the on-time date.** This one seems straight-forward until you start thinking about gaming the system. If the original due date is Oct 16 and it looks like that date will be missed, then if you change the date to Oct 18th, the order may, all of the sudden, look like it is on-time. You want to prevent that from happening. However, you may need to consider legitimate changes to the due date. For example, if the customer originally requests a due date of Oct 16, but then realizes that they don't have room for

the shipment or a need for it until Oct 23, then you may want to use the Oct 23rd date. Some companies have started to use two measures for the on-time date. They use both the **original date** and whatever the **adjusted date**. If the measure against the original date is far from the that of the adjusted date, you may find additional inefficiencies in understanding why the original date is changed frequently.

- **When do you calculate the measure.** When you measure can impact your service level measure. For example, if you only measure service level on items that have shipped, it is easy to temporarily game the system by continuing to delay orders that are already late. That is, once you ship the late order, it will show up late. But, if you delay it, it does not show up in the measure. You can get around this by counting all orders once they've past their due date. The other reason for determining when you measure the service levels is to wait for the order to arrive at your customer and to get confirmation that it arrived and that it arrived damage free-- that can take several days.

## Best Practices Using the Measures

As you can see with the discussion above, there is no perfect measure- each one has its benefits and limitations. And, you can see that there are many nuances to how you calculate the measure.

Here are some best practice tips we've gathered over the years:

1. **Keep in mind how you will make money by measuring service levels.** You don't want to measure service levels for the sake of measuring. Most times the goal comes down to making more money. Use this objective to figure out which measure you want to use. For example, if one of your major customers requires that you hit 95% OTIF to keep the business, then you use that measure. If you think you can gain a competitive position by having better service, you may have flexibility to choose the metric and how you best calculate it.
2. **Don't think you have to have one measure.** It can be a mistake, and lead to endless arguments to have to come down to one measure. Instead, like we discussed above with the original due date *versus* the adjusted due date, it might make sense to track both. Likewise, you may want to track Unit Fill Rate and OTIF to see how different the measures are with just a difference of ship date *versus* arrival date. Likewise, you might want to measure OTIF with giving yourself partial credit and without. In general, you will find the multiple measures can give you a richer picture of what is happening.
3. **If you need a global measure to compare, keep it simple.** If you need to use the measure to compare your different sites or business units, it is best to keep the measure relatively simple. All sites may not have the same ability to measure complicated service calculations (like Perfect Order). So, pick a measure or two that all sites will be able to support.
4. **Watch out for Service Measures that Contradict Other Groups.** This is an issue with any measurement system- improving one measure may hurt another measure. In the

example here, if you are trying to maximize OTIF, it is important that you keep a complete mix of your products in inventory so you are ready to ship. If your manufacturing organization is trying to maximize utilization, they may focus on long production runs and not be willing to change the line to make another product-- even if you need it to provide good service.

## Tracking Root Causes

The goal of tracking the service level is to improve it and make sure you continue to meet it. Just tracking and reporting on the measure will provide some of the benefit (this is the Hawthorne effect). Knowing that it is being measured, people will be more conscious of making sure orders go out on time. This effect can be significant.

However, to dramatically improve your service level, to get last bit of improvement (to go from 93% to 99%), or to make sure your changes are long-lasting, you need to use the service measures to help understand the root cause of the problem. Then, you can fix that root cause.

For example, if you track both original data and the adjusted date and you notice a big difference, you will realize that you have a problem with the date changing. Now, you want to dig into that issue to see if the problem is with your team calling the customer to change the date, whether the customer requests an unrealistic date and it needs to be changed, whether you are changing the date based on when you schedule the truck and so on. Each one of these issues has a potentially different solution.

You can do more with the data you are collecting. You will have information on all your orders-- both the good ones and the ones where you had a service failure. You can then use new machine learning algorithms to sort through this data to look for potential patterns for problems. For example, do orders with more than 10 items tend to be late? Do orders that are right at the cut-off between one and two day shipping (~400-500 miles) tend to be late? When you pick orders too close to the cut-off time, do you increase the chance of being late? And, so. The algorithms can analyze the root cause of the problems and allow you to address those root causes. This is opposed to the standard way of manually going through the service failures and labeling the failures by hand. The manual method is too time consuming and introduces too much bias.