Keysight Technologies
Too Much Calibration?

White Paper

Abstract

"Hey Wendle, our cost of sales is increasing, and my boss is riding me hard to reduce it! That calibration budget you submitted is "killing" me. It is just too much calibration. Get back to me tomorrow with a better proposal will you? " "Oh, and by the way could you look at our warranty failures and get that under control? We are up 5% from a week ago." As with another common euphemism, calibration, confidence, and expense all rolls down hill.

KEYSIGHT TECHNOLOGIES

Unlocking Measurement Insights

Introduction

High confidence in the test system directly translates to high confidence in the DUT test results and "high" costs buying that confidence. The opposite is true as well. Low confidence in the test system is a guarantee of low confidence in the DUT test results, but a great way to trim costs. Between the "too much" and the "not enough" is the elusive balance point between cost and confidence. This paper will explain the variables that are at work in the world of calibration, how they can be used to find that balance point, and ultimately answer those most important questions. "Will the product work?" and "How much is it going to cost me to know?" It will also provide a common denominator in the languages used by the metrologists and manufacturing.

Webster's Dictionary has always been a favorite of mine. He worked night and day to define our language to a "T", no pun intended. Then we create nuances around each word to match our needs in verbal communication. One would think that any words connected with the world of metrology would be meticulously defined and universally used and understood. That however is not the case. Here is some metrological and manufacturing vocabulary as described in Webster's.

Author: Todd B. Wendle
Keysight Technologies, Inc.

Accuracy: "Noun, freedom from mistake or error : correctness 2. conformity to truth or to a standard or model : exactness b : degree of conformity of a measure to a standard or a true value."

Calibrate: "Verb, to standardize (as a measuring instrument) by determining the deviation from a standard so as to ascertain the proper correction factors."

Now we look at a couple of words from the ma nufacturing test side of the fence.

Adequate: "Adjective sufficient for a specific requirement, especially: barely sufficient or satisfactory."

Confidence: "The quality or state of being certain: certitude."

So, right there in Webster's is the answer to the question "Too much calibration?" The question changes to a statement:

"Compare to a known standard a measuring instrument with barely sufficient accuracy as to have adequate confidence that the product being tested will perform as specified in the marketplace all at a reasonable cost."

Filled with subjectivity, this definition is doomed. The words mean different things to different people. My accuracy may not be near as accurate as yours. Your adequate may not meet my needs. And a consumer's confidence may not match with that of the manufacturer.

In order to gain that consumer confidence, the device performance has to be measured in such a manner as to provide "proof" (there is another one of those words) the results of the testing are accurate enough. "Accurate enough" has a lot of implications. It implies the device has to work correctly, is ready for use, will meet customer expectations, and will be at the lowest possible cost. And to make it more interesting, there is always the metrology request for 95% confidence levels. It just sounds good. And that is the basis for this paper. "Accurate enough" is a very difficult unit of measure. 95% confidence levels are sometimes expensive. How on earth can a metrologist survive where there is no "accurate enough" measurement device at a guaranteed 95% confidence? The real answer lies hidden within the performance characteristics of the testing equipment and the use of a unit of measure everyone understands, money.

How DOES the manufacturer "know" the testing equipment is providing the degree of certainty needed to have the device work correctly and meet, or hopefully, exceed customer expectations all at a reasonable cost?

Performance characteristics of a test instrument are defined by two primary variables. The two are measurement uncertainty of the test system and the measurement error as the measurement results drift over time. Note there are also uncertainties associated with the device under test (DUT). While these uncertainties are important, they are out of scope for this paper.

In order to establish the contributions of both variables, they must be separated and defined. For the purposes of this paper the first variable is defined to be the inherent uncertainties defined by the individual components of the test system. Also involved in the calculation of uncertainties are the connections between test instruments, fixtures used to hold the DUT and combination effects. Every piece of test equipment and test system has some degree of uncertainty. This uncertainty is dependent on many factors. As you look at the following graphic, the measurement error component has NOT been introduced. This is done later in the paper.

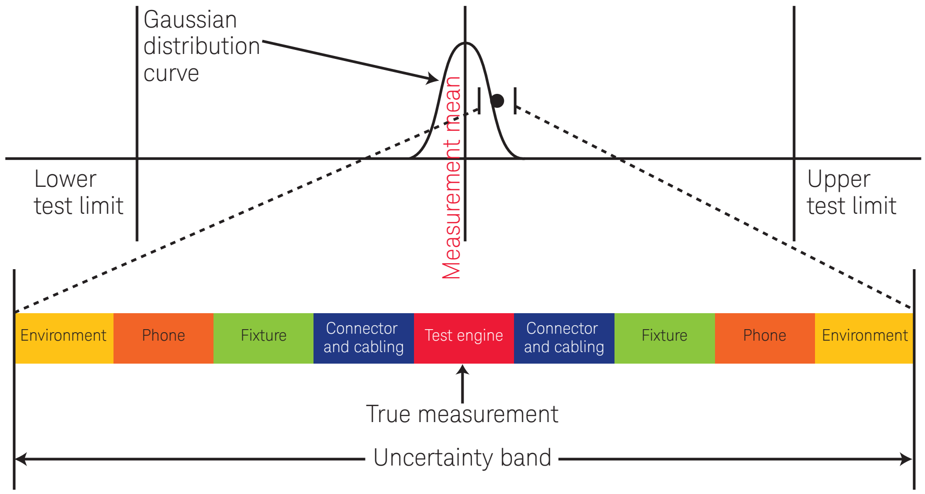

Figure 1. Measurement uncertainty

Figure 1 is an example of the composite uncertainty band in the testing of cellular phones. In this example, as in most test environments, the test is only run once and the single result determines pass or fail. If the DUT were to be tested more than once, the result has a high probability it will not be the same the second time. Nor would it be the same the third time. This variation in results is caused by measurement uncertainty and is illustrated with the colored band.

When that colored band of uncertainty crosses a test limit, See Figure 2., the uncertainty of the measurement result begins to be important. As the test result moves closer to the test limit, the importance of uncertainty continues to increase until there is a 50% chance the DUT is a true pass and a 50% chance that it is a true failure. Once the test result and its uncertainty band are beyond the test limit, the probability of being a true failure is 100%. Or is it?

This limited test determination creates two very important DUT populations. Those are false passes and false fails. An entire industry has been built around these two populations. These populations are at the center of the need for reduction of measurement uncertainty and correction of measurement error.

With measurement uncertainty defined, let's move on to the second variable, measurement error. Test instrument results can and do drift over time. The test result provided a year ago may very well be different if run today. In some instances drift can happen in a matter of minutes. Take the sides off of the test rack, and the ambient temperature drops over 10 °C in a few minutes, causing the spectrum analyzer to drift by over 1 dB. For the purposes of this paper, this drift is defined as measurement error and the measurement of that error as calibration.

To best define measurement error, one more diagram is needed to further the definition and effects of measurement uncertainty.

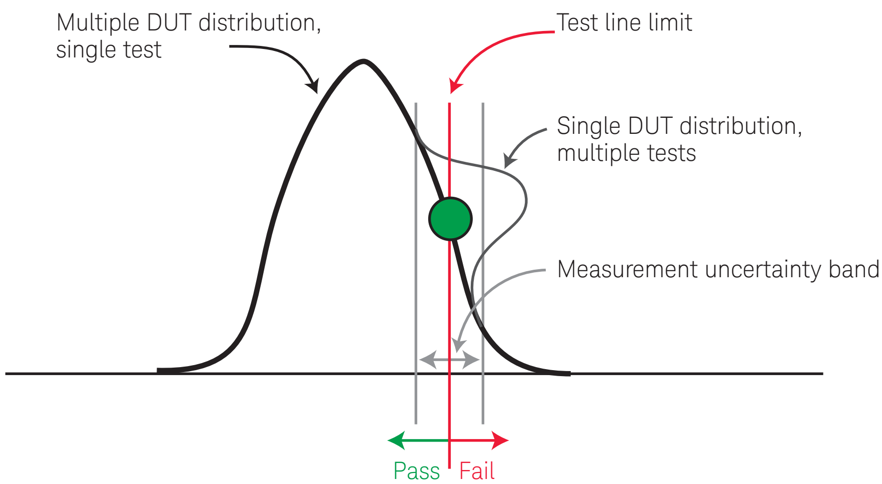

Figure 2. Test limit crossing multiple DUT single test normal distribution

Figure 2. is an illustration of a best-case situation where the test instrument output has been calibrated and then the results adjusted to the midpoint between its OEM specifications. The uncertainty band is the measurement uncertainty of the test instrument and other possible factors as described in Figure 1. With this situation, there is not a measurement error component in the uncertainty band. The green dot is a single DUT result measured once. The pink distribution curve is the distribution curve that is defined by running that single DUT multiple times. That distribution defines the measurement uncertainty associated with that particular test. Each single DUT test result has this measurement uncertainty distribution.

When there are DUT failures, the multiple DUTs single test distribution (this is the heavy black line in Figure 2) falls across one or both of the specification limits for that test. Around the specification limit are four populations; true passes, true failures, false passes, and false failures. In the purest sense, false passes are really true failures and false failures are really true passes. See Figure 3.

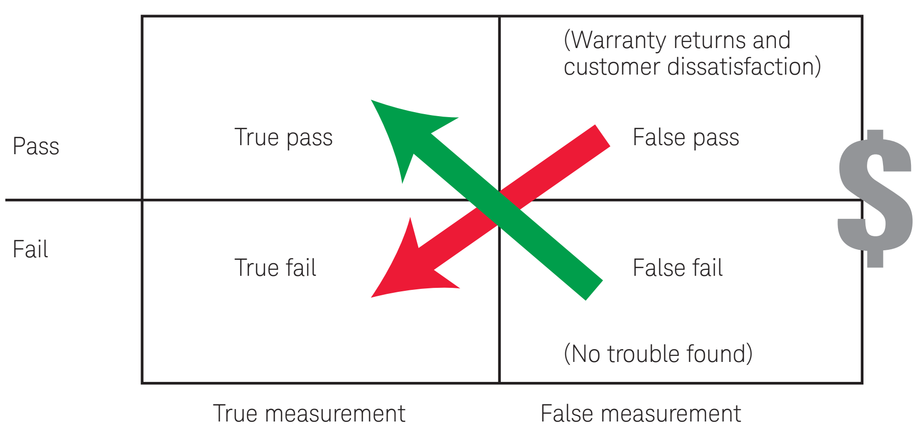

Figure 3. False passes and false fails

In that range of measurement uncertainty and measurement error, there is an ever-changing probability ranging from almost zero to 50% that the test result indicated is not correct. And if it is not correct, finding out which is correct will cost money. That is ONE thing that does have a 100% probability of occurring.

Figure 4 illustrates the test results of a single test where the test instrument's output for the particular test has drifted (over time) adding a measurement error component to the original uncertainty band. The single test result in Figure 2 had a 50-50 chance of being a true pass or a true fail. After instrument drift, the result will be a fail. This increase in total fails will have an increased population of false fails. The costs associated with this false fail population (all false fails attributable to the measurement error) create the demand for calibration.

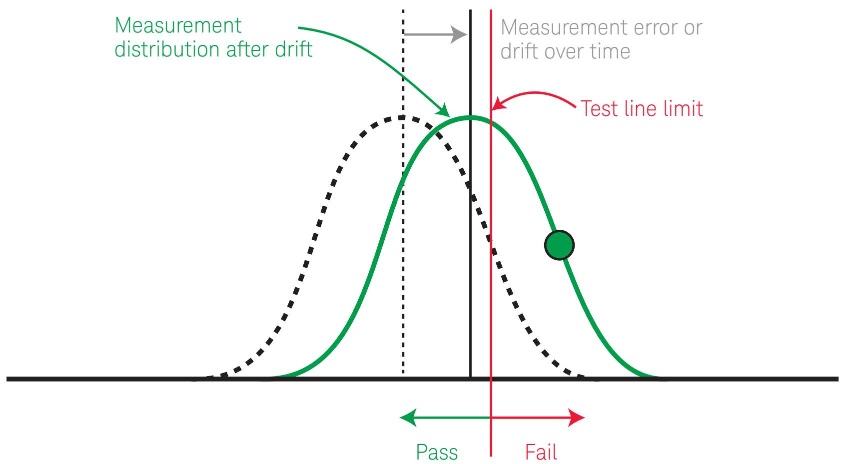

Figure 4. False fail drift

Now let's drift the other way. Figure 5 illustrates this event. In this case the measurement error drifts such that all the DUTs are now passing. As this is occurring, and since it is a gradual drift, production would be happy in that all their efforts to reduce failures is "working". Bosses all the way up the chain are happy, and back patting is available for all. But as with all "good" things, this comes to an unhappy end. These false passes are getting to the customer and they are sending them back under warranty for replacement. And maybe even more important is that repair operation databases are not directly coupled with warranty return databases. This means that it will take a long time, if ever, to recognize any connection between the two at an operational level. The failures are still "zero", but the warranty costs from this instrument continue to mount. Now you have a case where these mounting false pass costs create additional demand for calibration.

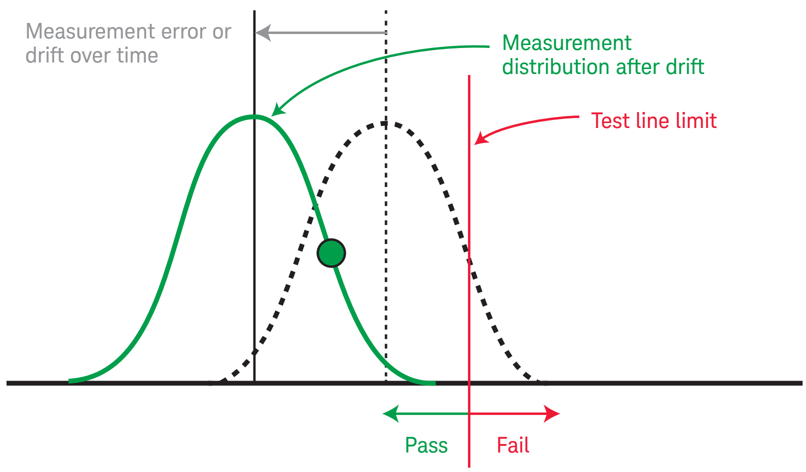

Figure 5. False pass drift

Instrument drift can have no effect, create additional false fails, and can create additional false passes. As the drift moves the DUT single measurement distribution past a test limit, the effects will show up as unexplained yield changes, good or bad, and after some period of time unexplained increases in warranty costs. Management sets the expectations for what are reasonable costs (yet another one of those words). In the manufacturing world, those costs levels are the result of the confidence level set for the test system. This can be compounded by multiple test systems on a single production line. The costs are created from true failures, false failures, and false passes. The confidence level of a test system will be improved for only one reason. That reason is money. Either make more revenue or cut more costs with expense reductions. There are no other reasons.

Confidence levels can be improved in three ways. One way is to reduce measurement error with calibration. The second way is to buy more expensive (read lower uncertainties) test instruments.

And the third way is to statistically account for the uncertainties by setting calculated “test limits” that are tighter than the actual test limits. This is commonly called guard banding. These guard bands “artificially” decrease the uncertainty band as illustrated in the previous Figures. As mentioned earlier, these second two ways will not be discussed in this paper.

Any confidence level improvement proposal will have an associated cost. To management, this cost must be justified with some type of return on investment analysis with a defensible and acceptable level of return. Reducing measurement error with calibration is no exception. With a robust calibration program, the calibrations will corrects problems before they ever occur thus making them difficult to quantify an ROI for calibration expenses. That does not mean it cannot be done however. As mentioned earlier, the heart of the matter is the bottom line. Answering the following three questions begins the process of knowing when there is “too much calibration” and too much impact to that bottom line.

- Does the test system produce DUT failures for any tests?
- Can warranty costs be attributed to a particular test?
- Can customer dissatisfaction be tied to false pass DUTs?

Let’s examine each of these three questions in more detail with respect to their effects on costs.

Question 1: Is the test system producing DUT failures?

There are only two answers to this question. Yes or no. First let's address "No". This means that 100% of the DUTs are passing this test. This can have several explanations.

- The test results are extremely stable over an extended period of time and there are no retesting, no reworking, and no warranty costs associated with this particular test. The DUTs has been designed with enough margins to insure no failures, and the combined effect of measurement uncertainty and measurement error has no effect on the validity of the test results.
- The test results have changed, either gradually, or all of a sudden, such that the yield has climbed to 100%. If there is no reasonable explanation for this change, there is a likelihood that there has been instrument drift or if the instrument has just come back from calibration or repair, the instrument is calibrated incorrectly and is passing true failures (Figure 5).
- The test limits have been artificially and unscientifically opened up to the point where all DUTs pass. Metrics drive behaviors and there is many an example of test limits being manipulated such that they give the desired test results.
- The test isn't done due to a bug in the test software.

Since improving yields is a good thing, not many people will complain about a reduction in DUT failures and its associated rework costs. This will only show up IF someone downstream of this test is impacted by the false results. If this happens within the factory walls, it will appear as rework. If it happens outside of the factory walls, it may very well appear as warranty returns costing much more.

Now let's look to the "Yes, this test is creating DUT failures" answer. There are three populations of failures; true failures, false failures (FF), and false passes (FP). True failures need to be reworked to work properly. No amount of calibration will affect these failures or the costs associated with them. However, the other two are directly related to calibration and uncertainty. FFs and FPs cost money. Please see Figure 3. They cost money in retesting (test time, opportunity cost of not testing new product), diagnosing false failures (labor and opportunity cost), and last false passes (downstream rework, warranty returns and customer dissatisfaction).

Question 2: Can warranty costs be attributed to a particular test?

This question begins the investigation of the false pass population. Note this FP population does NOT include any DUTs that fail with end user use. Those, at the time of test, were true passes.

If the answer is no, and there have been no changes in yields, false passes are not a significant portion of the DUT FF and FP Failure costs. If the answer is no, but there has been a change in yields it may still be worthwhile to figure out why.

If the answer is yes, measurement uncertainty and measurement error are creating false passes. These can follow two paths. False pass DUTs can show up as a failure further downstream where it would show up as rework in a different area at a typically higher cost, because the unit will be further assembled. This is important in that not attributing the rework to a specific cause, the opportunity for improvement is lost. The second and much more costly path is the false pass DUT that reaches the customer where it is determined to be a failure. It is sent back and becomes a warranty return with all the costs associated with a return. Whether rework or warranty, this is a direct cost from a false pass. A common rule of thumb is that a false pass costs a company 10X that of a false fail!

Question 3: Can customer dissatisfaction be tied to false pass DUTs?

This question is important. If the experience of a failure is unpleasant enough, the customer will ask for their money back, and more than likely tell friends and associates about the poor quality product just purchased. This is difficult to quantify, but it needs to be considered when evaluating the need for calibration and reducing false passes.

It is important to remember that not all false fail and false pass costs are included in the financial analysis for calibration needs. True failures most certainly have a cost, but are not attributable to uncertainty and measurement error. And further, it is important to separate DUT FF and FP measurement uncertainty costs from those associated with measurement error.

In this next section, we begin the process of quantifying the effect of measurement error. With some statistically valid data collection, a competent statistician can describe the distributions and the number of false passes and false failures to expect. The next step is getting a thorough understanding of the retest/rework processes and associated costs. The third step is a thorough understanding of costs of warranty returns and if possible, although more difficult, estimate the cost of customer dissatisfaction. Armed with this data, the projected measurement error DUT FF and FP costs can be calculated. At the very least, valid estimations can be made.

There are potentially four cost components to calculate. They are false passes that reach the end customer, false passes that are discovered downstream in the process, false failure costs, and customer dissatisfaction costs.

The first is false passes that reach the end user. This cost is calculated from statistical analysis of false passes multiplied by the average cost of a warranty return.

The second false pass population are FPs that are discovered downstream, and can be linked to being a FP from upstream, the FP costs for this subset is the rework costs associated with identifying the failure at a downstream rework process. Remember that a FP is really a true failure. It is only those additional costs in identifying the true failure that are use for this analysis.

The third potential cost component is false fails attributable to measurement error. If there is a retest built into the test software, the DUT FF cost is number of retests that pass the second time times the first time retest costs. The second part of this are those DUTs that fail twice in the production test system, but subsequently are diagnosed and retesting as false fails through some type of rework process. There are examples where parts are removed and replaced just to make sure the DUT is “fixed”. This portion of the cost is the number of FFs times the cost of diagnosing, retesting true passes, and unnecessarily replaced parts. Figure 6 is the process flow for the above described production environment with a population of false failures.

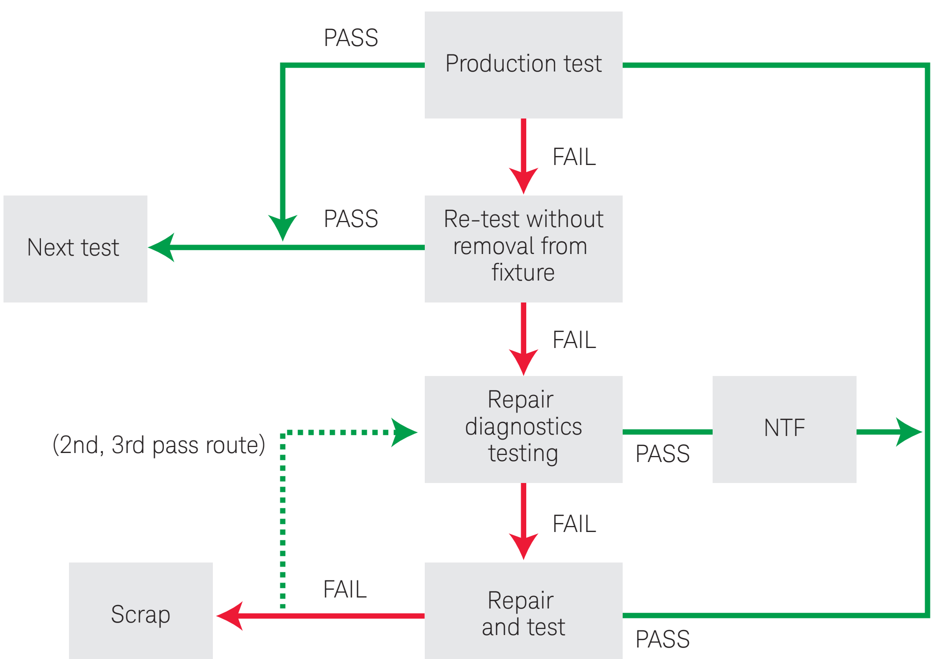

Figure 6. Typical false failure flow

Remember it is important to separate the DUT FF FP cost of measurement uncertainty from that attributable to measurement error. That can be done two ways. With historic data, the DUT FF FP costs can be graphed. At this point they include costs from both measurement uncertainty and measurement error. Using the OEM specifications of the test instrument and some statistic work including the test limit, the cost of measurement uncertainty DUT FF FP costs can be calculated. Subtracting this from the overall DUT FF FP cost provides the DUT FF FP costs from measurement error.

The other method would be to calculate the costs right after a calibration and then before the next calibration. The delta between the two is the cost attributable to measurement error. (Assuming the measurement uncertainty does not change over the time period.)

There is a fourth potential for cost. It is the customer dissatisfaction component. It is subjective and difficult to quantify. It is very real however and all that is needed is to stop and remember a moment in your past where a product you were using failed. If a defensible estimate can be made, by all means include it in the analysis.

Components of the DUT FF FP costs associated with the existing test instrument should be now fairly clear. What about the costs of doing calibrations? What is not easy is figuring out what level of calibration is needed, the point at which a calibration is needed, and even if to calibrate the instrument at all. And to really make this ugly, there are wide variations in the cost of calibrations. What supplier do I choose? Is the expense of calibration worth it?

The “level” (both type and frequency) of calibration services plays a very important role in establishing an adequate confidence. The key word is adequate. Metrologists, like statisticians, “error” on the side of high confidence low risk exacting results. This level of rigor collides head on with the manufacturing world of DUT testing. There is no perfect measurement. There is a level of error inherent in all measurements. That error can be minimized, but is inversely proportional to the dollars spent to reduce it. But it is never eliminated.

Each test result distribution that crosses a test limit will have a measurement error and measurement uncertainty component. Understanding the impact of instrument drift is key in determining a particular calibration. Beyond that, a calibration with stated test accuracy ratios or documented uncertainties might very well be a necessity if the measurement uncertainty component for the test instrument is to be quantified. The next step is to insure the calibration has the thoroughness needed to understand the test instrument’s performance over the ranges required to test the DUT. All calibration services are not created equal. Test points are eliminated, and in some cases full tests are eliminated. Calibration providers typically offer a range of calibration services from minimal tests to a fully accredited calibration.

And last, but not least, there may be certain calibration standards required by your customers. ISO 17025 is emerging as such a standard. It will be more than regular calibrations, but it provides additional documentation and a higher degree of assurance that the calibration you are receiving meets certain technical requirements.

It was pointed out that metrics can and do drive behaviors. Someone is responsible for yields, warranty costs, downtime, etc. That someone is the business operations manager. To him or her, high confidence low risk means higher costs (more expensive equipment and calibration expenses), greater downtime (more calibration time), and lower yields (more DUT failures from increased confidence levels). These are the three metrics they are usually responsible for improving, not making worse. Increasing calibration is not typically a popular decision. However, understanding these effects and knowing the financial impact of measurement error will help deal with these issues at the right level in the organization. It will provide solid justification for reducing measurement error with particular calibrations and calibration cycles.

Now we have the two sides of the general equation.

Calibration expense = Summation of daily FF+FP costs over calibration cycle + (FF + FP reduction savings over calibration cycle). One side is the financial impact of measurement error and on the other side the cost of the calibrations themselves. A very important point is to keep things equal. The time period to measure the costs on either side of the equation must be the same. Calibration costs typically happen once per year while the DUT failure costs are daily. They are very seldom compared let alone put into the same equation.

There are really three specific situations with the general equation. They are simple enough, but the work to determine the values of the actual variables is not.

When f(t)CE [1] = f(t)SumFFPE [2], calibration is balanced, but evaluate the potential cost savings from increasing calibration expenses (to reduce FFFP measurement error costs)

When f(t)CE [1] < f(t)SumFFPE [2], evaluate increasing level and or frequency of calibrations

When f(t)/CE > f(t)SumFFPE [2], calibration may be too thorough and or too frequent, but evaluate as insurance (How many FPs are you willing to have with a reduced calibration level or calibration cycle?)

f(t) = measured time period
CE = calibration expenses
FFPE = DUT false fail, false pass expenses as a direct result of measurement error

1. Usually a single occurrence event.
2. Usually an ever-increasing amount related directly to amount of drift and relationship to FF and FP expenses.

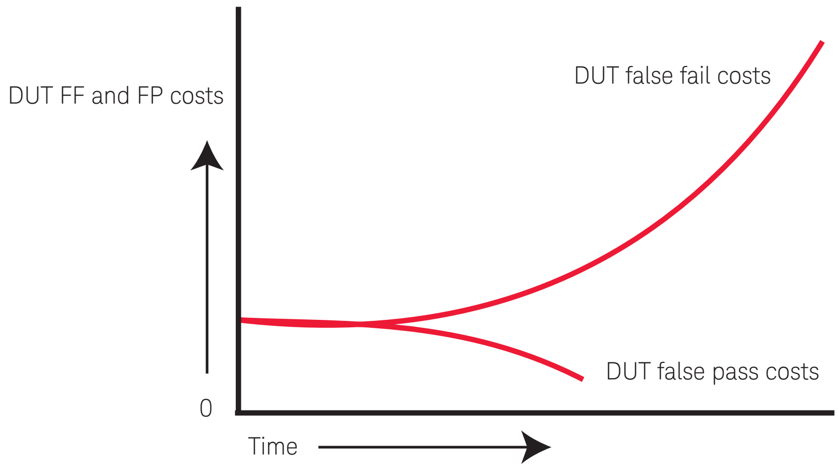

Figure 7. Crossing the line

Looking at this graphically helps demonstrate the points. The dollar impact changes over time. Figure 7 is an example of a test where the FF FP costs are stable. The test equipment performance begins to drift and the multiple DUT single test distribution moves towards a test limit. The first indication of this is decrease in yields. The next indication is an increase in DUT FF costs and a decrease in FP costs. The combined FF FP costs begin to climb.

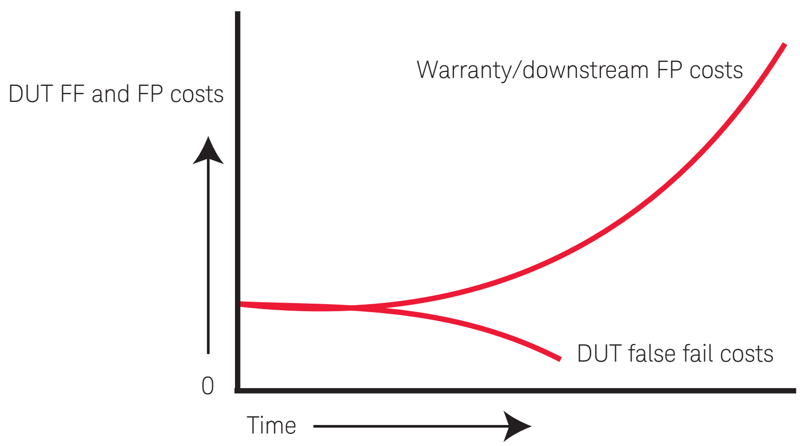

Figure 8. Improving yields?

Figure 8 is the other possibility. Again there is a steady state of DUT FF and FP costs. It is in “control”. Then the test instrument begins to drift away from the test limit such that the yield improves, but it is only improving due to increasing false passes. DUT FF costs go down, but the warranty costs and customer dissatisfaction begin to climb. To make matters worse, there can be a time lag in seeing the warranty costs and as mentioned earlier, there may very well be no connection made between the two seemingly unrelated events. Again, the combined FF FP costs begin to climb.

Both Figure 7 and 8 illustrate what can happen over time. Given the ever-worsening results, something will be done eventually. But there is another possibility. There is a balance point between calibration expenses and the delta between DUT failure FF and FP costs associated with measurement error calculated over the calibration cycle time.

This balance is the area of adequate calibration.

Depending on specific ROI analysis, higher- level calibrations or more frequent calibrations may provide much higher returns.

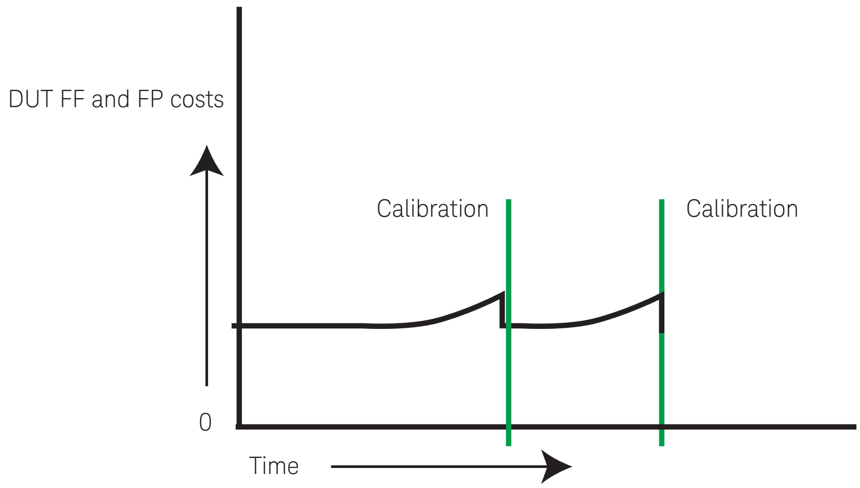

Figure 9. Calibration effects

The step function graphics shown in Figure 9 represents the financial impact when doing periodic calibrations. It is the return for your calibration investment. Even when there are no failures and it appears there is no immediate return on your calibration dollar, calibration can provide value. Looking at Figure 7, it is easy to see how if the test instrument should drift or malfunction, there is a high likelihood there will be rework costs or warranty costs. Calibration helps insures this situation will not happen. Without a rework/warranty benchmark to calculate calibration returns, calibration becomes insurance. Is the "cost" of the insurance worth the potential loss without it? Graphing your particular situation can go a long way in visually seeing what calibration can buy you.

One important note to consider is that as the volume and or rework/warranty costs per DUTs go up, the calibration ROI can go up very quickly.

In conclusion, there are many variables at work in the world of calibration. Each situation is directly impacted by the specific variables and their values. Measurement error causes false passes and false fails. These two populations are expensive. They affect the bottom line. Separating measurement uncertainty and measurement error false results cost opens the door to calibration justification. Calibration can and will improve your profits, your expenses, and your customers satisfaction!

And for that ever-present manager, one final comment.

"Hey Mr. Manager, that new calibration program report with full financials is on your desk. "

"Oh, and by the way our warranty failures are down 10%. That new calibration program is working just as planned."

Keysight Services

www.keysight.com/find/KeysightServices
Flexible service solutions to minimize downtime and reduce the lifetime cost of ownership.

Keysight Infoline

www.keysight.com/find/service
Keysight's insight to best in class information management. Free access to your Keysight equipment company reports and e-library.

KEYSIGHT TECHNOLOGIES
Unlocking Measurement Insights

Published in USA, May 5, 2015
5991-1311EN
www.keysight.com