

A young woman with dark hair and glasses is smiling while holding a smartphone. She is wearing a white t-shirt. The background is a blurred office environment with a desk, a pen holder, and a computer monitor. The image is overlaid with a blue and white graphic design consisting of curved shapes.

# Effectively Benchmarking your Phishing Awareness Program

**SANS**

**SECURITY  
AWARENESS**

# When orchestrating any security awareness program,

it is only natural to try to gauge an organization's overall security position against a jury of one's peers, so to speak – that is, measure success against comparable organizations and programs. The growing need to benchmark these programs has created a bit of a flurry over benchmarking methodologies of late within the phishing arena, particularly given the absence of industry standards on everything from program maturity to awareness tools and phishing templates.

This eBook addresses benchmarking and examines what to keep in mind when using external or internal benchmarks to assess your phishing program. We will cover which variables to consider while creating benchmarks, and what to look for when comparing your results across industries. We'll also cover global and internal benchmarking as well as concepts such as tiering.

# 01



## What Is Benchmarking?

What is a benchmark, anyway? And how many variables need to be similar for the benchmark to be valid, or even make sense? According to Bernard Marr & Co., benchmarks are reference points that can be used to compare your performance against the performance of others, typically across internal business entities or external competitors. Business processes, procedures, and performance analytics are compared taking into account the best practices and statistics from similar organizations.

While there are several types of benchmarks, phishing falls under performance benchmarking, which uses performance metrics. This type of benchmarking involves comparing not only against like companies or competitors in similar industries, but also making global comparisons regardless of industry. It can even involve internal benchmarking across your own company, such as comparing different departments, regions, or business units.

Phishing simulations are used by many companies across all industries as a key cyber training tactic to teach people how to better identify and stop phishing attacks where adversaries use deception to gather sensitive and personal information. Phishing simulations generally measure the undesired action rate (click rate) and the report rate (number of reports generated).

Most Security Awareness teams are interested in how their phishing program metrics compare to similar businesses, and it can be tempting to place too much emphasis on performance benchmarking. While identifying areas of improvement based on comparable industries is important, it is also important to track multiple variables using phishing simulations and results as a primary driver.

Some variables include:

- The report rate
- The history of reports
- The difficulty of the simulation
- The expected time to report
- The top 10 most common phishing tactics
- When the simulation was completed (day/time)
- The ease of reporting
- The availability of reporting channels and awareness
- "Time in training" (for example, the number of employees who have completed the simulation)
- Other relevant metrics
- The overall Security Awareness score

# Program Maturity

When benchmarking, it is important to consider your phishing program in relation to your larger security awareness efforts and the maturity of your program as a whole. For example, if your phishing program lands on the higher end of the **SANS Security Awareness Maturity Model**<sup>®</sup> – that is, you have previously promoted cyber awareness tactics and experienced long-term sustainment and culture change – then your phishing assessment results, even with a similar phishing simulation, might be very different from those of a similar organization with a less mature program still in the compliance-focused phase of the model. The workforce under the compliance-focused phase will have had less time to use teachable moments and safe cyber awareness tactics.

For more information about this maturity model, visit <https://go.sans.org/lp-ebook-maturity-model>.



SANS Security Awareness Maturity Model<sup>®</sup>



Non-existent

Compliance-focused

Promoting Awareness & Behavior Change

Long-term Sustainment & Culture Change

Strategic Metrics Framework

# Simulation Difficulty

Another variable to consider is the difficulty of the simulations. If the simulations across each business or organization are identical or very comparable, the statistics will be more valid. If the simulations you are using for benchmarking vary in difficulty, the benchmark metric most likely will be skewed. This is why more mature phishing programs often use a concept called tiering, which digs deeper into the nature of the simulations and indicators used.

While benchmarking has its place and can be used to gain support from leadership, care must be taken to ensure that the benchmarking is of “apples to apples” wherever possible. Benchmarking is a worthwhile assessment tool only if you have the capacity to consider all the factors and variables. Otherwise too much emphasis on benchmark statistics – instead of on more impactful metrics such as the report and click rates – can drive the wrong behavior.

Together, program maturity and strategic tiering will provide the baseline measurements necessary to perform meaningful benchmarking. The next chapter in this eBook will explore tiering in more detail.



# 02

## How to Effectively Use Phishing Benchmarks to Assess Your Security Awareness Program: The SANS Tiering Model

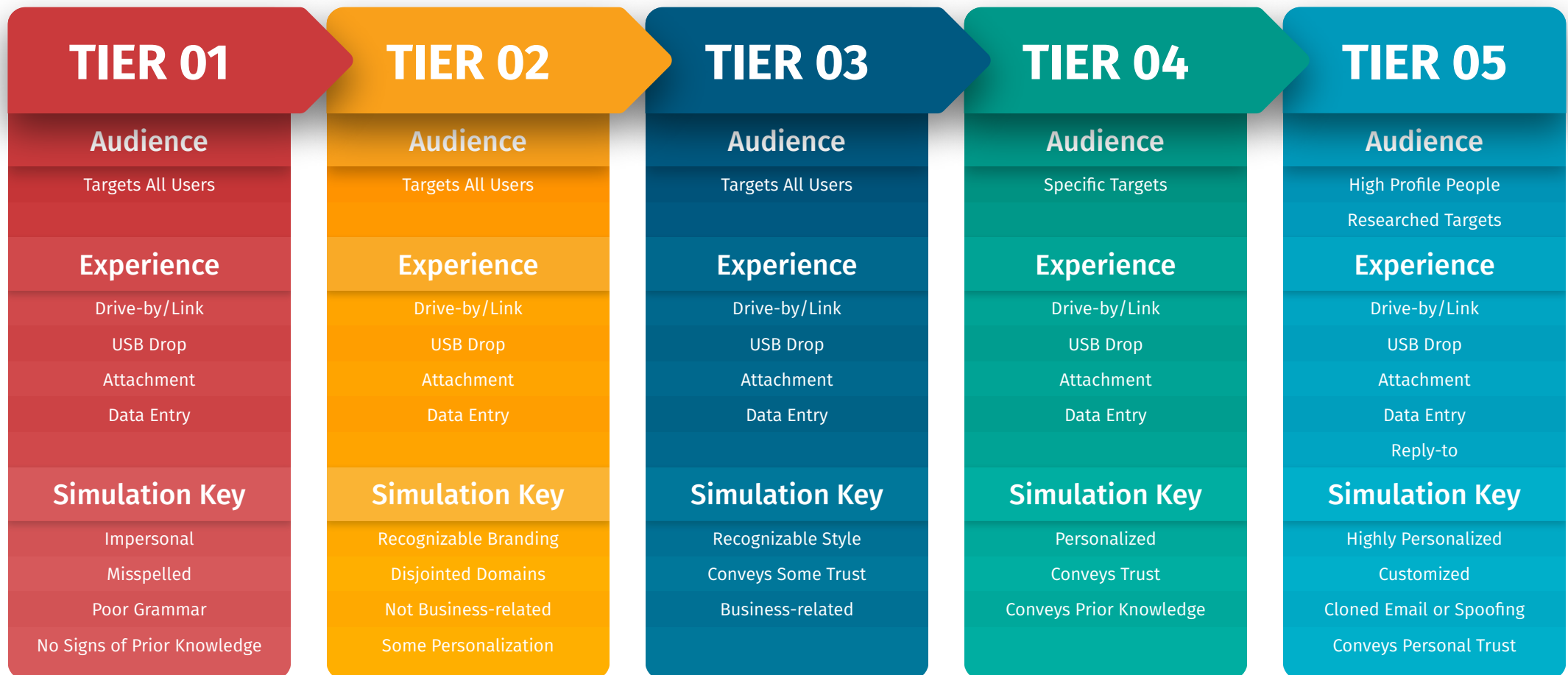
Chapter 1 introduced the concept of benchmarking in your phishing programs, the importance of selecting the correct metrics to track, and how those metrics relate to your progress on your organization's maturity model. This chapter will address tiering, which allows for a more valuable assessment of your simulation data analysis, or an "apples to apples" comparison. The tiering concept identifies the difficulty of your simulations, how many indicators are involved, and how hard it is to identify and recognize these indicators as a phish.

The SANS tiering model has five levels. Each tier has three specific areas: audience, experience, and simulation keys. While the audience and experience might vary from tier to tier, the simulation keys are what truly separate the tiers.

As shown in the diagram below, a **Tier 1** simulation would be very “spammy” in nature, easy to recognize as a phish, and impersonal. As you move up in the tiering, the indicators are more difficult to spot, with brands or memo styling with which the workforce may be

familiar. This could be a non-business-related shipping notice in the **Tier 2** block, or a Zoom business-related simulation within **Tier 3**.

**Tier 4** includes personalization and some trust and prior knowledge, such as program status. **Tier 5** simulations become very personalized and highly targeted to specific individuals and executive leaders. Assessing these simulations requires additional resources, and such assessments typically are separate events outside most simulation schedules and benchmark data.





As can be seen in the sample **Tier 1** simulation below, the indicators are easy to pinpoint, and should be quickly identifiable by a well-trained workforce.

The image shows a simulated email interface with the following content and callouts:

- Callout 1:** "There is no specific addressee in the 'To:' line" points to the "To: Ecard User" line.
- Callout 2:** "There is no trust, or even knowledge of, sender" points to the sender information "Ecard Team <eCard@daily-winner.net>".
- Callout 3:** "Spelling or grammatical errors are present" points to the underlined text "[This was so funny!](#)".
- Callout 4:** "Lacks recognizable logo or internal company format" points to the circular logo containing the letters "ET".

The email body text includes: "A friend just sent you an ecard from e-hugs-online.com", "You can view it by clicking heer:", "[This was so funny!](#)", "Using our new tracking feature, you can now view all the ecards received by you in the last 30 days.", and "[Your ecard is going to be with us for the next 30 days.](#)".

Tiering is especially important when benchmarking with external entities. Benchmarking is often easier said than done, since so many variables impact the validity of the results. Many of these variables, such as the demographics of the workforce or the simulation schedule, are easy to identify. However, identifying the level of difficulty of the phish is harder to determine because it can be dependent on vendor tools and how simulations are labeled based on difficulty, if identified at all.

With such a wide range of organizations now using phishing assessments as an awareness tool, it has become easier to identify similar industries within a given vertical. It can be tempting to further narrow this down by size and even workforce maturity, though, typically, specifics and similarities with regard to the actual simulation data are hard to compare. Using a tiering model provides the ability to better “bucket” simulations, ensuring that the benchmark data are as viable as possible.

For example, if an organization requires the assessment of its overall security position via-à-vis comparable organizations, it could look at

historical simulation assessments and select those that fit the selected tier. When each organization in an industry shares metrics within the given tier, the metrics are far more impactful than metrics across all tiers, or all difficulty levels.

The ideal situation involves collaboration with similar industries and cybersecurity programs – that is, sending the identical simulation to a represented sample size. If the workforce demographics are comparable, then the undesired action rate, along with the rate of employees reporting the phish, will provide a trustworthy and actionable benchmark if indicated.

Communication and cooperation across the external entities is crucial, but it may be difficult based on the typical workload of the security awareness officer and operational teams. Carving out the time and resources to execute phishing simulations, and establishing strategic priorities, could drive the team to rely on historical analytics with tiering considerations. In this situation, the team could rely on bucketing the simulations in a like tier, reducing the resource and scheduling impact.

## Internal Benchmarking

The final chapter in this eBook focuses on internal benchmarking, which involves comparing different entities within your company or enterprise. Typically, the goal of internal benchmarking is to gain a better understanding of how entities in your company are doing. With phishing simulations, you can analyze the results to determine which entities are responding well to phishing simulations, and which may need additional work in terms of awareness.

Internal benchmarking will be highly dependent on your company structure, the layout of your organization, departments, business units, regions, and teleworking, access to awareness and training materials, and even “time in band,” that is, how long employees have been in their current role.





At first, internal benchmarking may seem easier to tackle than external comparisons. It uses a common set of characteristics that most likely are similar across the company or can be planned accordingly, including:

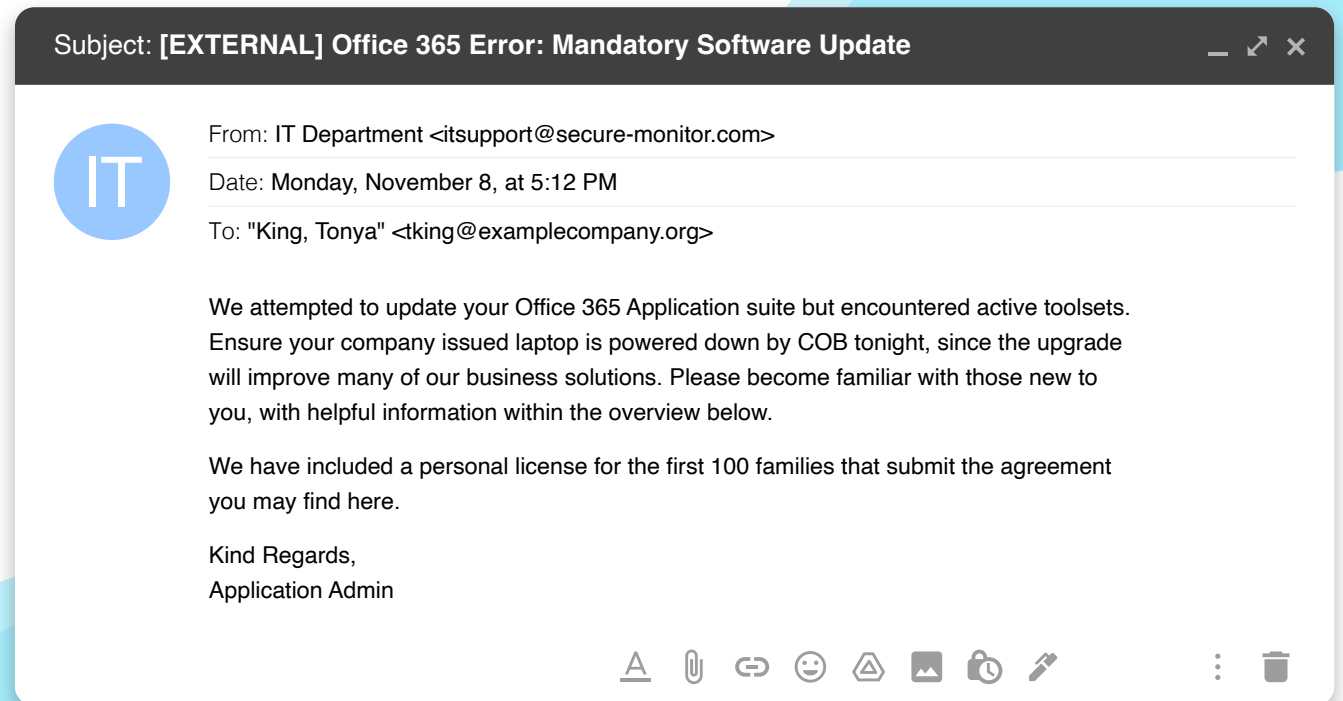
- Overall position within the **SANS Maturity Model®**
- Representative sample size of the workforce
- History and length of the phishing program
- When the simulation was sent (day/time based on location)
- Difficulty and experience of the simulation, and how many indicators are present (providing a lower tier and not based on roles)
- Ease of reporting and reporting options

However, other factors and tactical aspects may require thought, strategic planning, and historical trending (if available), including:

- How relative the simulation might be to participants
- The organization and distribution of departments and roles across the business
- The availability, variety, and distribution of training and awareness materials
- Leadership, management, and stakeholder engagement

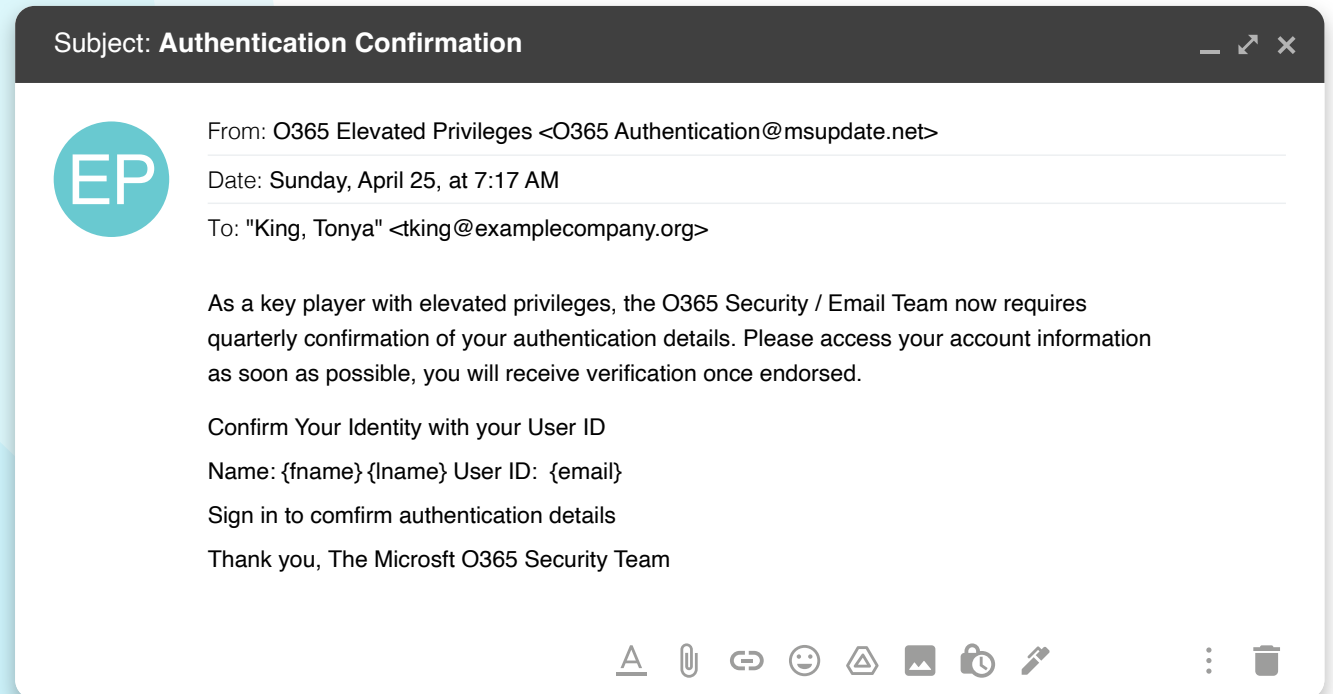
As with external benchmarking against other organizations, comparing “apples to apples” is important when validating the results of internal benchmarking. The previous chapter addressed tiering in order to ensure the validity of benchmarks. When using benchmarks as an analytic tool to evaluate your phishing program, the analytics should look at the difficulty of the phish, how many indicators present themselves, and how hard it is to distinguish, isolate, and identify something as a phish.


Based on the tiering concept, we can see that while one simulation might fit all employees within the company, a more advanced simulation may not. A **Tier 1** or **2** simulation would be appropriate for the entire workforce, but typically not targeted to a specific business, role, or skill. **Tier 3** or **4** simulations may not be appropriate for the entire workforce. These more sophisticated options may only apply to specific groups or roles across the organization and are more business-related, while often conveying trust or prior knowledge. The example below is appropriate for employees with company-issued laptops and could result in a higher undesired action rate than those employees without company-issued devices. It would also hopefully result in a higher report rate.



In all cases, it is imperative for the workforce to have had exposure to security awareness and training materials. This would not be appropriate if there were many new employees who were just issued a new device and had yet to receive awareness materials.

While the prior example was based on a situation, the phish shown below is aimed at a specific role. The employees targeted here hold elevated and advanced credentials or authorizations.





Another factor to consider when measuring a phishing program is the diversification and distribution of awareness and training materials. If you have a wide range of demographics, ensure that your teachable moments are meaningful and understandable across your workforce. We all learn and consume information in different ways – some by print, others through videos and interactive options. Also consider, if applicable, the different physical locations and cultures of your benchmark groups. On-site availability of awareness and training resources might be more diverse than for a totally remote workforce or even a hybrid setup. Your benchmark data will be more valid if you've incorporated diverse awareness and training methodologies.

How your stakeholder, management, and leadership teams handle security awareness is also a factor to consider when assessing internal benchmarking data. The hope is that all levels of managers and stakeholders – including Human Resources, Communications, or IT – recognize the importance of improving and enhancing cybersecurity and social engineering skills. However, the background of leaders, organizational priorities, and limited resources can impact that focus. A remote field supervisor might feel differently about proper cyber behavior than an IT leader, and dedicated program managers faced with tight schedule deadlines may not promote security awareness as much as they should.

Internal benchmarking is a valuable tool: comparing metrics allows you to leverage those findings to entice the workforce into being more aware while developing the skills to recognize and report a phish. Some options to compare include:

- Corporate-wide simulations that apply to everyone, **Tier 1 or 2**
- Similar roles across business entities, including IT administration, administrative assistants, developers, human resource business partners , and even new employees
- Comparable departments such as Finance and Accounting, Research and Development, and Compliance
- Parallel management levels, such as directors or associates

Your security teams can also provide suggestions on target groups to benchmark, as those teams often have the resources to pinpoint the most vulnerable employees. In addition, your historical data derived from your phishing program can uncover weak departments, teams, and other areas that would benefit from additional training opportunities.

# CONCLUSION

## How to Benchmark Your Security Awareness Program

Benchmarking is an important undertaking for many facets of an organization. Your cybersecurity awareness program is no exception, as this eBook has shown. This is especially true if your program is not delivering expected results. Benchmarks allow you to more easily identify areas to reduce human risk to manageable levels. While it is tempting to quickly compare your program to those of competitor organizations based on intuition or limited research, the pragmatic, step-by-step approach outlined in this eBook will drive more meaningful results and a more successful phishing program.

For more information, please visit [sans.org/awareness](https://sans.org/awareness).