amazon

Manage Your Experiments Best Practice Guide

Run better experiments and understand your results.

What you'll learn in this guide

Version 1.1 | Last updated October 3, 2023

Introduction

What is Manage Your Experiments (MYE)?

Amazon customers use product detail pages to decide what to buy. Registered brand owners can enhance these product detail pages by adding rich content, diving into product features, or sharing their brand story. Used effectively, product title, main images, bullet points, product descriptions, and A+ Content can help drive sales and increase conversation rates. But how do you know what content converts?

As a brand owner, you can run A/B tests (also known as split tests) on your listing content with Manage Your Experiments. Experiments let you compare two versions of content against each other so you can see which performs better.

Why should I use MYE and what are the benefits of experimentation?

In 2022, brands reported up to a 25% increase in sales with the use of MYE. If you are a registered brand and meet the minimum traffic eligibility criteria, you can run experiments across hundreds of ASINs simultaneously across multiple marketplaces. Experimentation can yield the following benefits for your business:

Improve Conversion Rates
Understand the impact of optimizing content by running experiments on high impact ASINs. Brands improved conversion by 15% using MYE.

Increased Sales
A/B testing is the simplest and most effective means to determine what content converts to intent to purchase. Brands have reported up to 25% increase in sales.

Showcase Your Brand
By experimenting with A+ Content you can stand out against the competition with optimized content, tell your story to customers, and help customers make the best shopping decisions.

How to build powerful experiments

1. What makes for a good experiment?

Building experiments with a specific hypothesis by determining what factors impact a customer throughout their shopping journey is the first step toward knowing what excites customers about the product, and what should be tested for further validation.

A good experiment starts with a scientifically guided hypothesis. What is a Hypothesis? A hypothesis is a prediction about the outcome or result of your experiment with an ultimate goal to change customer behavior. It may not guarantee a winning result and your hypothesis may be proved or disproved. Either way, an experiment is a learning opportunity and can help build another experiment based on the learning of the previous one. A good hypothesis takes into account past customer behavior and turns it in to a prediction that can positively impact the customer journey. In situations where such data is unavailable and you would like to conduct experiments on a new product without historical data, an initial A/B test will still provide guidance on what content resonates with customers better for improved customer engagement.

A hypothesis has 3 components:

“Changing [X] by adding/deleting [Y] will result in [Z].”

[X]: The original version of your Title/Image/Content.

[Y]: The element of the original version of your Title/Image/Content that you wish to change or remove, elements or words you would like to add to the original version to make it more appealing to customers.

[Z]: The change in customer behavior you wish to see or the result you are going to track from the experiment. Example, conversion rate, click thru rate, units sold etc.

Examples of a good hypothesis:
Changing [my title] to add the [brand name in the beginning of the title] will [increase sales by 6%.]

Removing [the picture of the dog] from [my image] will [increase conversion rate by 15% by appealing to customers above the age of 12].

Examples of a bad hypothesis:
This is not an effective hypothesis: “My conversion rate will increase as a result of this experiment.” You can improve this hypothesis by considering how to design your experiment, ask yourself, “What elements can I modify or change to achieve my desired results?”

TIP

Don’t assume your experiment is complete once you’ve hit statistical significance. Consider what you learned and determine what you can iterate on for your next test.

2. When to experiment?

Good A/B tests are driven by a large sample size and optimum duration to capture the nuances of customer behavior and external trends that impact conversion rate and purchase intent. Consider the seasonal appeal of your product, does the content need to be updated once a year, or twice a year to meet the in-trend demand and customer behavior? For example, promoting a portable speaker as a gift item might resonate more with customers during the holidays, whereas in the summer time, it can be connected to camping or other lifestyle use cases.

We recommend you experiment all the time, everywhere!

> If you double the number of experiments you do per year, you're going to double your inventiveness.
>
> 
>
> **-Jeff Bezos**
> Founder of Amazon

Market trends and customer behavior are volatile and change constantly. It's good practice to run experiments to identify these changing trends and surface the most up to date content that resonates with your customers "right now." Before scheduling an experiment, consider the following tips:

- Schedule an experiment with a promotion, sponsored Ad, or new product launches to drive more traffic and get faster and better results. Onkata, a brand that has run over 500+ experiments on MYE, reported +250% increase in sales by publishing the results of an experiment when it was run in conjunction with a sponsored ad or promotion.
- Avoid making too many changes in one experiment, as this may lead to unclear findings on what drove the difference in results. For example, for a piece of apparel, if you changed the title by adding the dimensions (S/M/L), target audience (adults/teens/kids), and a specific feature (color/texture) or details of the product, it would be unclear as to which element influenced the result.
- Consider what external factors might skew the results of your test. For example, if your product is a sports jersey, it might be a good idea to test your product during events like the Super Bowl, as customers are more inclined to view your product and you will get more traffic on your ASIN. Testing your product off-season may not get you enough insights if it's designed for a niche segment. You can try testing a couple of days before the start of the season and during the season to capture maximum traffic for your experiment.

3. Interpreting the results of an experiment

Experiment results are updated once a week until the experiment ends. The results detail how well each version of your content performed with customers, along with the probability of how much better the winning content performed over the alternative. You can access your results by clicking on the experiment name in the experiments dashboard.

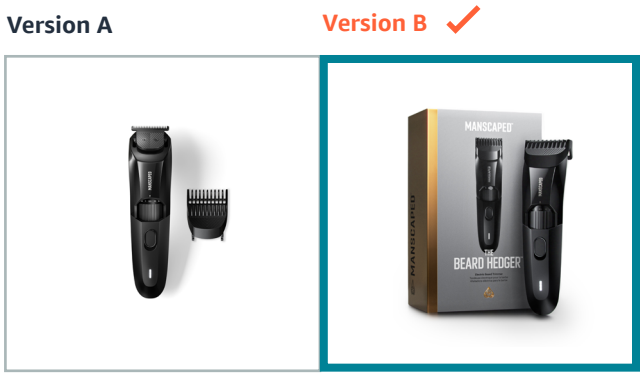
Based on the data collected during the experiment, we calculate a range of possible impacts of publishing each piece of content. Results are aggregated for all enrolled ASINs in an experiment. We provide a few kinds of results:

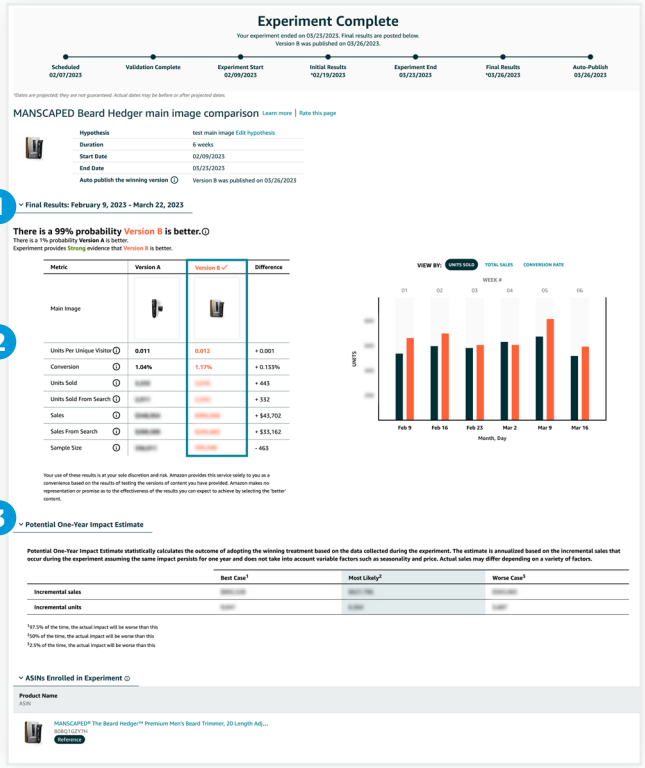
1. **Probability that one version of content is better.** For example, if we say there is a 75% probability that Version A is better it means that 75% of the possible impacts we calculated show a likely positive units/sales lift from publishing Version A.

2. **For each version of content, we show: Units, sales, conversion rate, units sold per unique visitor, and sample size assigned to that version.**

   - The conversion rate is the percentage of unique visitors in the experiment who saw the content and made a purchase.
   - The sample size is the count of unique visitors who saw each version of content.
   - The units per unique visitor is units divided by sample size.

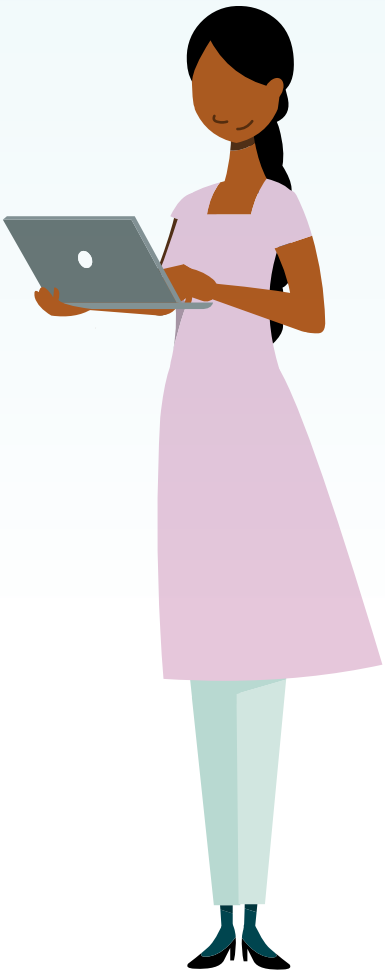
3. **Projected one-year impact.** This section is only populated for completed experiments. It's calculated based on the average daily sales increase of the winning content and multiplied by 365 to simulate the long-term impact. This is an estimate that doesn't take into account seasonality, price changes, or other factors that would affect your business in the real world. It's provided for informational purposes only and incremental benefits are not guaranteed.





Recommended general tips

Make sure your experimental content is different from existing content: If your content is too similar, it is less likely to affect customer behavior, and you may not be able to confidently determine a winner.



Tips for experiment setups

- Let your experiment run for the entire duration: Even if early results are promising, they can be misleading. If you end an experiment early, you may increase the chance of overestimating the impact of your experiment or even selecting the wrong version.
- Use the pre-selected optimum settings to begin your experiment as soon as your validation is complete.
- Use the pre-selected Experiment to Significance if unsure of ideal duration. This feature will conclude your experiment automatically when the data reaches significance.
- Use the pre-selected Auto-Publish feature to automatically publish the winning content once test is complete.

Tips for title, product description, and bullet point experiments:

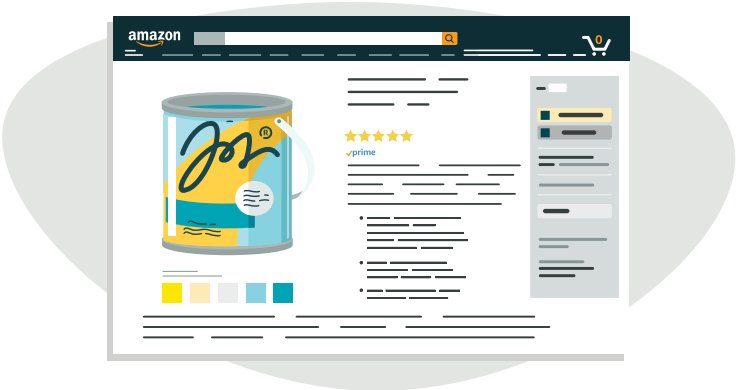
- Try adding or removing the brand name: Including a brand name in a title may have a positive or negative impact based on customer perception of the brand. While it may be useful to include the brand name for established and well-known brands, it might have the opposite effect on brands that are just starting out or aren't as well known. It could be useful to test both scenarios to understand the value and perception of your brand. Additionally, it's important to know that customer perceptions evolve overtime. As more customers use your products, the brand perception might improve. It can be beneficial to experiment on the same hypothesis from time to time to see how customer perception has evolved and display content accordingly.
- Try adding benefits of the product in the title: Adding benefits to the title may help introduce customers to your differentiating factor and help improve click-through rate. While it may not guarantee a purchase, it could pique interest in the product.
- Try adding or removing product specifications: Specifications that could enhance the value of a product may be valuable information in a title. Market research could help guide you toward popular features or specifications within your customer segment. These could also change quickly and market research should be continued. For an Amazon option through Seller Central, refer to Product Opportunity Explorer.
- Shorten product titles: Try reducing the title length to under 100 characters to reduce noise and encourage more customers to visit your detail page. Also, ensure you emphasize differentiating aspects that might appeal to customers in the title.

Tips for A+ Content experiments:

- Experiment on the exact same ASIN: When you identify two versions of A+ Content to test in an experiment, make sure both versions are applied to the exact same set of ASINs using the A+ Content Manager. Otherwise, you won't be able to set up your experiment.
- Introduce different A+ Content modules to your detail pages: Replace standard image modules with dual images with text module or add comparison charts.
- Try adding or removing the brand name: Including a brand name in the A+ Content may have a positive or negative impact depending on customer perception of the brand.

Tips for image experiments:

- Add creative product imagery to stand out from the competition.
- Add high-resolution images in your content and refresh your existing images.
- Use more lifestyle imagery over standard product imagery.
- Highlight unique-to-brand feature sets to stand out from the competition and other offerings.



Frequently asked questions

Who is included in the experiment?

Experiments are based on individual customer accounts. During an experiment, each customer account that sees your content is considered part of the experiment. Customers are randomly assigned to view one version of content and they will see that content persistently for the duration of the experiment regardless of device type or other factors, as long as the customer can be identified. Visits to your page where a customer cannot be identified are not included in the sample size. Each session only counts unique visits per customer. If a customer visits the page 10 times in the same day , we only count 1 visit per day.

What does it mean when version B is 90% better than version A?

There is 90% chance that version B is better than version A. This does not represent the magnitude of difference between unit sales for A and B. For example, version B has 50 more units sold than version A in 90% of the cases.

What does it mean when the experiment is inconclusive?

An experiment can end with results that are inconclusive, or results that show a low confidence that one version of content is better than another. However, these results can still be valuable.

Here are some reasons why an experiment may have inconclusive results:

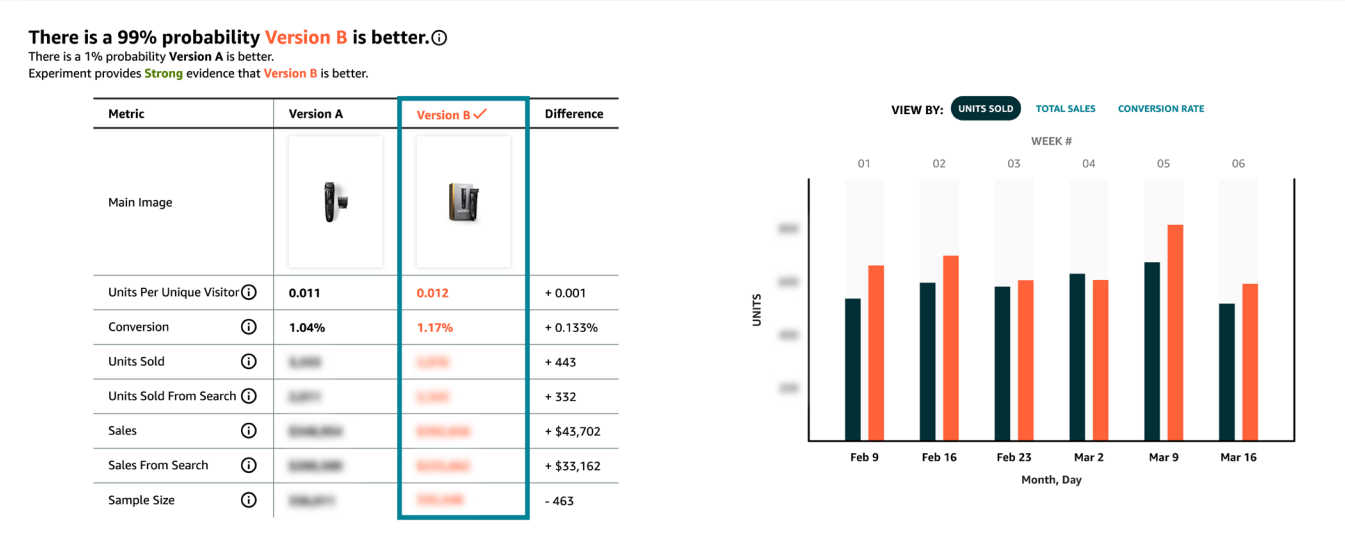
- The change you made to your content was too small to significantly change customer behavior
- There wasn't enough traffic to determine the winning content with high confidence
- The two versions of content you tested were similarly effective in driving sales

Refer to your experiment hypothesis when trying to make sense of inconclusive results. For example, depending on what you changed, an inconclusive result can tell you that a certain type of content isn't worth investing in because it doesn't affect customer behavior. Or, it can tell you that two ways of merchandising your product are similarly effective. You can run additional experiments to confirm what you've learned from your earlier tests.

How is the winning content determined?

We use a Bayesian approach to analyze experiment results. This means we construct a probability distribution based on a model as well as the actual results of the experiment. We report the mean effect size (in terms of change in units) as well as the 95% confidence interval (also known as a credible interval) of the posterior probability distribution, which is updated weekly during the experiment based on all experiment data collected since the start. The confidence of a winning treatment is the percentage of outcomes in the probability distribution that show a positive unit sales impact.

How are A/B test results determined?



An A/B test has three main factors that determine the results of the test:

(i) **Sample Size:** MYE has established a minimum threshold per month for experimentation because sample size is the most important determining factor to ensure validity of the results of your A/B test. The bigger the sample size, the smaller the margin of error. If the test gets stopped before each variation reaches its required number of visitors, the test will conclude with results that are less reliable. Your test should reach the required sample size per variation for the results to be valid.

(ii) **Significance level:** Statistical significance indicates that the differences observed between the two versions are large enough to conclude the experiment, in this case 66% or better. An experiment will reach statistical significance only when enough data has been accumulated by customers viewing the content associated with a given experiment. Each time a piece of content is viewed, we log which content was seen per customer and whether or not the product was purchased in order to calculate statistical significance.

(iii) **Duration of test**: The ideal duration for an A/B test depends on the product, its business life cycle, and the amount of traffic it receives. Every product has its own low and high selling season as well as a customer traffic pattern in a given week. For example, a product that usually sells on weekends is likely to have lower traffic and conversion during weekdays.
To get valid test data, you should run your test across low and high seasons, as well as weekdays and weekends. Additionally, the lower the traffic, the longer you will have to run the test. We recommend running an A/B test using Experiment to Significance to ensure you're getting the most reliable results without the inconsistencies due to fluctuating traffic patterns.

Start experimenting now!