Experimentation Hub

08/01/2025 12:25 pm EDT

The Experimentation Hub allows you to review and manage your campaigns' A/B tests and Holdout Groups in one centralized location within Bluecore. This will save you time from needing to go into each campaign to review the A/B Test or Holdout data. Now, you can see every campaign with an active, paused, or completed test and view its latest results in the Experimentation Hub.

Experimentation Hub optimizes your time spent on testing by letting you know when a test is ready to declare a winner. It will save you time calculating results - you don't have to worry about calculating whether a result is statistically significant anymore.

> ⓘ Your experiments will still need to be set up individually at the campaign level.

A/B Tests

A/B Tests allow you to test different elements of a campaign to make data-driven decisions about which aspects of a campaign are performing best. Use A/B Tests to compare two or more subject lines, product recommendations, offers, or more.

This is a look at the Experimentation Hub:

Filters

There are filters at the top of the dashboard you can use to adjust which tests to view.

1. Active Status

   - Filters the A/B Tests into three possible categories:

     1. Active - Tests that are currently running

     2. Complete - Tests that have finished and have results that can be viewed, with a full backlog of all previous tests

     3. Paused - Tests that aren't yet complete, but have been paused

2. Analysis Metric

   - The currently selected metric for the analysis of this test:

     1. Conversion Rate - The number of customers who purchased, divided by the number of customers who received an email.

     2. Unique CTR - The number of customers who clicked, divided by number of customers who received an email.

3. Significance Threshold (P value)

   - Adjust the level of confidence of the result of the test. A lower P value means more confidence in the result of the test. For example, a P value of 0.02 means there's a 2% chance of the

result occurring due to random chance alone, instead of via customer actions.

4. Is Significant?

   ○ Filtering just the statistically significant tests will only show tests that meet the significance threshold in the previous filter. For example, a test with a 0.15 P value would be considered significant if the significance threshold is 0.2, but not if the significance threshold is 0.1.

5. Channel

   ○ Email or SMS

6. Campaign

   ○ Narrow the dashboard to show data for only specific campaigns.

7. Journey

   ○ Narrow the dashboard to show data for only specific Journeys.

8. The percentage of A/B Tests currently achieving statistical significance, based on the filters above. Narrowing the selection of tests will change this percentage.

9. Number of tests ready for action, based on the filters above. Narrowing the selection of tests will change this number.

In the Experimentation Hub, "ready for action" means one of two things:
- At least one variant in the test has reached a statistically significant result
- The test has been active for at least 30 days, and you should assess whether the test should continue

Under the Ready for Action list, you can view campaigns that need attention and meet one of the above criteria.

All Tests View

Campaign Name [1]	Start Date [2]	Name [3]	Type [4]	Test Metric [5]	Description [6]	State [7]	Variant Allocation [8]	Days Active [9]	Delivered [10]
	2024-12-18		⌀		⌀	Active	50%	26	64,511
	2024-09-23		⌀		⌀	Active	50%	112	162,764
	2024-09-23		⌀		⌀	Active	50%	112	69,792
	2024-12-18		⌀		⌀	Active	50%	26	65,523
	2024-09-19		⌀		⌀	Active	50%	116	78,955
	2024-12-03		⌀		⌀	Active	50%	42	51,860
	2024-09-23		⌀		⌀	Active	50%	112	205,904
	2024-09-19		⌀		⌀	Active	50%	116	18,077
	2024-10-08		⌀		⌀	Active	50%	97	69,546
	2024-09-23		⌀		⌀	Active	50%	112	21,146
	2024-09-23		⌀		⌀	Active	50%	112	99,826
	2024-09-23		⌀		⌀	Active	50%	112	70,949
	2024-09-23		⌀		⌀	Active	50%	112	83,265
	2024-12-18		⌀		⌀	Active	50%	26	49,632
	2024-12-18		⌀		⌀	Active	50%	26	128,523

Under All Tests, you can see all tests, see how long they've been active, whether they've reached statistical significance, and the latest results of the tests.

1. Campaign Name

   a. The name of the campaign in Bluecore.

2. Start Date

   a. Date the A/B Test began.

3. Name
    a. The name of the A/B Test.

4. Type
    a. The type of test chosen when setting up the campaign
        i. Incentives
        ii. Recommendations
        iii. Creative
        iv. Subject Line/Copy
        v. Personalization
        vi. Custom

5. Test Metric
    a. Winning criteria selected when setting up the campaign
        i. Conversion Rate
        ii. Click Through Rate

6. Description
    a. Additional information added when setting up the campaign

7. State
    a. Displays if a campaign is Active, Paused, or Complete.

8. Variant Allocation
    a. Percentage of the delivered emails are the test variant.

9. Days Active
    a. Days the A/B Test has been/was active.

10. Delivered
    a. Test variant email delivered by the campaign.

11. Total Delivered
    a. Total emails delivered by the campaign across all variants.

12. Analysis Status
    a. Displays the latest status of the test, such as "Significance reached," or if it still may need

more time.

13. Analysis Metric

    a. Determined by one of the filters chosen at the top of the page.

14. Variant ID

    a. Displays "Variant B."

15. Variant Name:

    a. Displays the name of the test variant entered in Bluecore when setting up the A/B test.

16. Test Variant Performance

    a. Performance of the test variant, based on the Success Metric chosen.

17. Variant A Performance

    a. Performance of Variant A, based on the Success Metric chosen.

18. Difference vs Variant A

    a. How much better or worse the test variant is than Variant A, as a percentage. Green is positive, red is negative. A green result means the test variant performed better than Variant A.

19. Is Significant?

    a. Displays whether the result of the test is statistically significant.

20. P value

    a. The confidence level of the result, as expressed as a decimal. A lower P value means more confidence in the result of the test.

21. Action

    a. Any potential actions will appear here as links. Clicking on "Review Results & Select Winner" will redirect you to the Bluecore campaign page, where you can see the same results and choose a winner of the A/B Test.
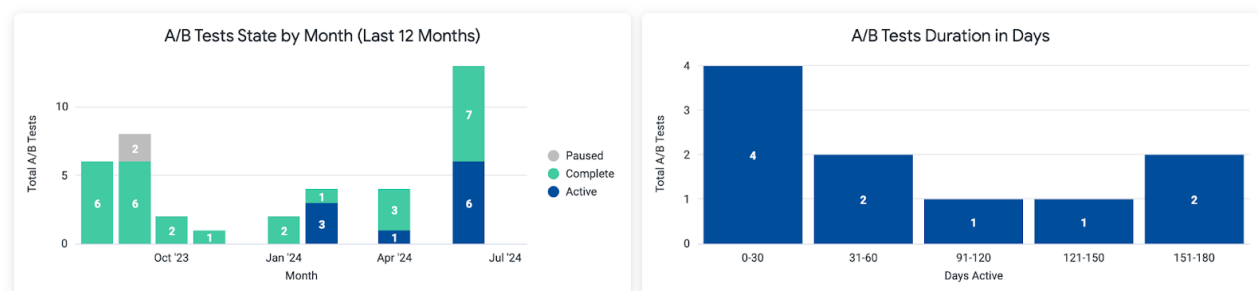
There are two insights available to learn more about the current tests - A/B tests by month, and by active days.

These can help demonstrate how many A/B tests have been historically run and for how long.

To get the best use out of the Experimentation Hub, check it about once a week to see if any new tests have reached statistical significance.
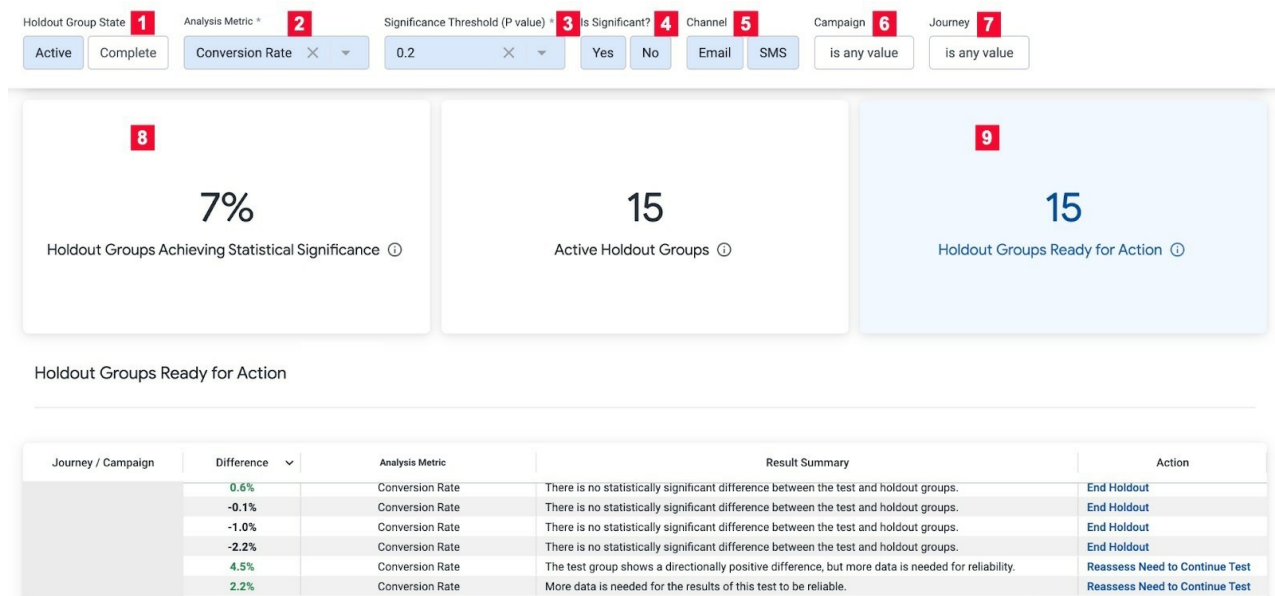
Test Variants

In Experimentation Hub, Variant B is always shown as the test, no matter what you call your variants when setting up your A/B Test. We recommend making Variant A your control and using Variant B (and/or variants C, D, and E) as the test.

This way, you can see how the test variant performed against the control in the Difference vs Variant A column - a green, positive number means Variant B (the test variant) performed better than Variant A. A red, negative number indicates Variant A performed better than Variant B.

Holdout Groups

Holdout Groups allow you to keep a percentage of eligible recipients from seeing your campaign, letting you know how effective the campaign is when compared to not seeing it.

This is a look at the Experimentation Hub:

Holdout Groups Ready for Action

Journey / Campaign	Difference ⌄	Analysis Metric	Result Summary	Action
	0.6%	Conversion Rate	There is no statistically significant difference between the test and holdout groups.	End Holdout
	-0.1%	Conversion Rate	There is no statistically significant difference between the test and holdout groups.	End Holdout
	-1.0%	Conversion Rate	There is no statistically significant difference between the test and holdout groups.	End Holdout
	-2.2%	Conversion Rate	There is no statistically significant difference between the test and holdout groups.	End Holdout
	4.5%	Conversion Rate	The test group shows a directionally positive difference, but more data is needed for reliability.	Reassess Need to Continue Test
	2.2%	Conversion Rate	More data is needed for the results of this test to be reliable.	Reassess Need to Continue Test

Filters

There are filters at the top of the dashboard you can use to adjust which tests to view.

1. Holdout Group State
    a. Filters the tests into two categories:
        i. Active
            i. Tests that are currently running

        ii. Complete
            i. Tests that have finished and have results that can be viewed, with a full backlog of all previous tests

2. Analysis Metric

    a. The currently selected metric for the analysis of this test:

        i. AOV Per Customer

            i. Tracks the average order value of customers in the test group versus those in the holdout group

        ii. Conversion Rate

            i. Tracks the percentage of customers who made a purchase in the test group compared to the holdout group

        iii. Cross Sell Conversion Rate

            i. Tracks the percentage of customers in the test group who purchased additional products in other categories versus those in the holdout group

        iv. Number of Orders per Customer

            i. Tracks the number of orders per customer in the test group versus the holdout group

        v. Revenue per Customer

            i. Tracks the average revenue per customer in the test group versus the holdout group

3. Significance Threshold (P value)

    a. This will adjust the level of confidence of the result of the test. A lower P value means more confidence in the result of the test. For example, a P value of 0.02 means there's a 2% chance of the result occurring due to random chance alone, instead of via customer actions.

4. Is Significant?

    a. Filtering just the statistically significant tests will only show tests that meet the significance threshold in the previous filter. For example, a test with a 0.15 P value would be considered significant if the significance threshold is 0.2, but not if the significance threshold is 0.1.

5. Channel

    a. Either Email or SMS, depending on if you want to specify a channel.

6. Campaign

    a. Narrow the dashboard to show data for only specific campaigns

7. Journey

    a. Narrow the dashboard to show data for only specific Journeys

8. The percentage of Holdout Group tests currently achieving statistical significance, based on the filters above. Narrowing the selection of tests will change this percentage.

9. Number of tests ready for action, based on the filters above.

In the Experimentation Hub, "ready for action" means one of two things:
- At least one variant in the test has reached a statistically significant result

- The test has been active for at least 90 days, and you should assess whether the test should continue. This is intentionally different from the A/B Tests using 30 days as a benchmark.

30 Days vs 90 Days

With Holdout Groups, more data is needed because it tracks the purchase behavior of the same customers over an extended period, requiring a larger sample size to account for factors like dropouts or the time needed to see some customers go through more than one buying cycle.

In contrast, A/B Tests compare outcomes between two groups at a single point in time, which typically needs less data to achieve statistically significant results.

Under the Ready for Action list, you can view campaigns that need attention and meet one of the above criteria.

All Tests View

Journey / Campaign [1]	Start Date [2]	End Date [3]		Test Metric [4]	Expected Lift [5]		Description [6]		Holdout Allocation [7]	State [8]	Days Active [9]	% of Min. Sample Size [10]
BC_abandon_browse_...	2024-10-18	∅	∅		∅	∅			5%	Active	87	100%
BC_abandon_search_4...	2024-10-18	∅	∅		∅	∅			5%	Active	87	100%
BC_abandon_search_3...	2024-10-18	∅	∅		∅	∅			5%	Active	87	100%
BC_abandon_category...	2024-10-23	∅	∅		∅	∅			5%	Active	82	100%
BC_abandon_search_2...	2024-10-18	∅	∅		∅	∅			5%	Active	87	100%
BC_abandon_browse_...	2024-10-18	∅			∅	∅			5%	Active	87	100%
BC_wishlist_1	BC_wi...	2024-09-27	∅			0%				5%	Active	108	8%
BC_abandon_checkou...	2024-09-27	∅			0%				5%	Active	108	4%
BC_reactivation_churn...	2024-10-01	∅			0%				5%	Active	104	17%
BC_abandon_category...	2024-10-18	∅	∅		∅	∅			5%	Active	87	8%
BC_post_purchase_on...	2024-12-17	∅	∅		∅	∅			10%	Active	27	0%

Under All Tests, you can see all tests, see how long they've been active, whether they've reached statistical significance, and the latest results of the tests.

1. Journey/Campaign

   a. Name of the Journey or campaign in Bluecore.

2. Start Date

   a. Date the Holdout Group test began.

3. End Date

   a. Date the Holdout Group test ended.

4. Test Metric

   a. Winning criteria selected when setting up the campaign.

      i. Conversion Rate

      ii. Revenue per Buyer

      iii. Revenue per Order

5. Expected Lift
   a. Percentage lift for the campaign when compared to the Holdout Group, selected when setting up the campaign.

6. Description
   a. Additional information added when setting up the campaign

7. Holdout Allocation
   a. What percentage of the campaign emails are/were being held back from delivery.

8. State
   a. Whether the campaign is Active, Paused, or Complete.

9. Days Active
   a. Days the Holdout Group test has been active.

10. % of Min. Sample Size
    a. How close the campaign is to reaching a minimum sample size for statistical significance. Some of the factors here include total participants and holdout allocation.

11. Participants
    a. Customers who did not get emails who otherwise would've.

12. Analysis Status
    a. Either "Minimum sample size reached," "Running for 90+ days," or "Running for 180+ days" depending on if the test has actionable results.

13. Analysis Metric
    a. Determined by one of the filters chosen at the top of the page.

14. Test Group Performance
    a. Performance of customers who received the campaign, based on the Success Metric chosen.

15. Holdout Group Performance
    a. Performance of customers who didn't receive the campaign, based on the Success Metric chosen.

16. Difference vs Holdout
    a. How much better or worse the test group is vs the Holdout Group, as a percentage. Green is positive, red is negative. A green result means the test group performed better than the Holdout Group.

17. Is Significant?

    a. Displays whether the result of the test is statistically significant.

18. P value

    a. The confidence level of the result, as expressed as a decimal. A lower P value means more confidence in the result of the test.

19. Action

    a. Any potential actions will appear here as links. Clicking on any links here will redirect you to the Bluecore campaign page, where you can see the same results and choose to continue the Holdout Group.

There are two insights available to learn more about the current tests - Holdout Group tests by month, and by active days. These can help demonstrate how many Holdout Group tests have been historically run and for how long.

To get the best use out of the Experimentation Hub, check it about once a week to see if any new tests have reached statistical significance.