Why Is The Control Video So Hard To Beat?

AUGUST 2020 UPDATE

High-performance creative is a rare thing for social advertising. In our experience, after spending over $3 billion dollars driving UA across Facebook and Google, usually only one out of 17 to 20 ads can beat the "best performing control" (the top ad). If a piece of creative doesn't outperform the best video, you lose money running it. Losers are killed quickly, and winners are scaled profitably.

The reality is, a vast majority of ads fail. The chart below shows the results of over 17,100 different ads. Spend is distributed based on ad performance. As you can see, out of those 17,000 ads, only a handful drove a majority of the profitable spend.

The high failure rate of most creative shapes creative strategy, budgets and ad testing methodology. If you can't test ads quickly and affordably, your campaign's financial performance is likely to suffer from a lot of non-converting spend. But testing alone isn't enough. You also must generate enough original creative concepts to fuel testing and uncover winners. Over the years, we've found that 16 out of 20 ads fail (5% to 17% success rate), you don't just need one new creative: You need 20 new original ideas or more to sustain performance and scale!

The high failure rate of most creative shapes creative strategy, budgets and ad testing methodology. If you can't test ads quickly and affordably, your campaign's financial performance is likely to suffer from a lot of non-converting spend.

But testing alone isn't enough. You also must generate enough original creative concepts to fuel testing and uncover winners. Over the years, we've found that 16 out of 20 ads fail (5% to 17% success rate), you don't just need one new creative: You need 20 new original ideas or more to sustain performance and scale!

And you need all that new creative fast becausecreative fatigues quickly. You may need 20 new creative concepts every month, or possibly even every week depending on your ad spend and how your title monetizes (IAA or IAP). The more spend you run through your account, the more likely it is that your ad's performancewill decline

So, we test a lot of creative. In fact, we produce and test more than 100,000 videosand images yearly for our clients!

But we don't just test a lot of new creative ideas. We also optimize our creative testing methodology. That might sound a little "meta," but it's essential for us to be validate and challenge our approach and results.

When we choose a winning ad out of a pack of competing ads, we'd like to know that we've made the right decision. When we kill most of our new concepts because they didn't test well, we react by moving on and pivoting the creative strategy based on results to try and find other ideas that may perform.

Because the outcomes of our tests have consequences - sometimes big consequences - we test our testing process. We question our testing methodology and the assumptions that shape it.

How We've Been Testing Creative Until Now

For the past few years, in an effort to streamline our Facebook and Google creativetesting and reduce non-converting spend, we've been testing new video concepts using IPM (Impressions Per Install) as the primary metric. For the record, using IPM is not the Face-book recommended best practice to allow ad sets to get out of the learning phase by gathering enough data to become statistically valid.

When testing creative we typically would test three to six videos along with a control video using Facebook's split test feature. We would show these ads to broad or 5-10% LALs (Lookalike) audiences, and restrict distribution to the Facebook newsfeed only, Android only and we'd use mobile app install bidding (MAI) to get about 100-250 installs.

If one of those new "challenger" ads beat the control video's IPM or came within 10%-15% of its performance, we would launch those potential new winning videos into the ad sets with the control video and let them fight it out to generate ROAS.

Unexpected Results

We've seen hints of what we're about to describe across numerous ad accounts and have confirmed with other 7-figure spending advertisers that they have seen the same thing. But for purposes of explanation, let's focus on one client of ours and how their ads performed in creative tests.

In November and December 2019, we produced +60 new video concepts for this client. All of them failed to beat the control video's IPM. This struck us as odd, and itwas statistically impossible. We expected to generate a new winner 5% of time or 1 out of 20 videos - so 3 winners. Since we felt confident in our creative ideas, we decided to look deeper into our testing methods.

Traditional testing methodology includes the idea of testing a testing system or an A/A test. A/A tests are like A/B tests, but instead of testing multiple creatives, you test the same creative in each "slot" of the test.

If your testing system/platform is working as expected, all "variations", should produce similar results assuming you get close to statistical significance. If your A/Atest results are very different, and the testing platform/methodology concludes that one variation or another significantly outperforms or underperforms compared to the other variations, there could be an issue with the testing method or quantity of data gathered.

Here's how we set up an A/A test to validate our non-standard approach to Facebook testing. The purpose of this test was to understand if Facebook maintains a creative history for the control and thus gives the control a performanceboost making it very difficult to beat - if you don't allow it to exit the learning phase and reach statistical relevance.

- We copied the control video four times and added one black pixel in different locations in each of the new "variations." This allowed us to run what would look like the same video to humans but would be different videos in the eyes of the testing platform. The goal was to get Facebook to assign new hash IDs for each cloned video and then test them all together and observe their IPMs.

- These are the ads we ran... except we didn't run the hotdog dog; I've replaced the actual ads with cute doges to avoid disclosing the advertiser's identity. IPMs for each ad in the far right of the image.
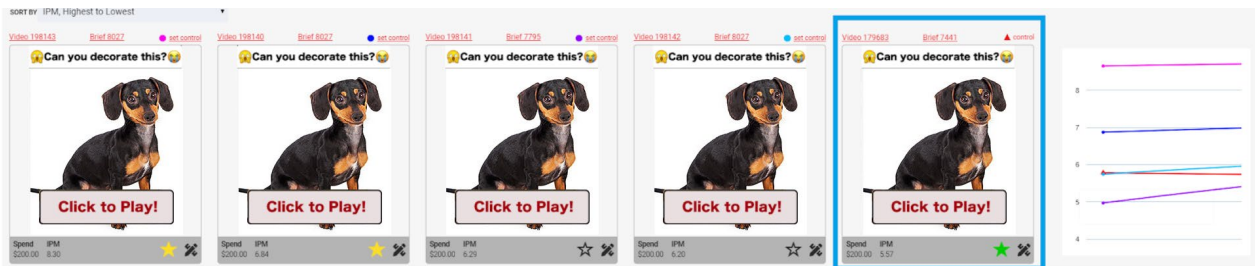
Things to note here:

- The far-right ad (in the blue square) is the control.
- All the other ads are clones of the control with one black pixel added.
- The far-left ad/clone outperformed the control by 149%. As described earlier, a difference like that shouldn't happen. If the platform was truly variation agnostic, BUT - to save money, we did not follow best practices to allow the ad set(s) to exit the learning phase.

  We ran this test for only 100 installs. Which is our standard operating procedure for creative testing.

  Once we completed our first test to 100 installs, we paused the campaign to analyze the results. Then we turned the campaign back on to scale up to 500 installs in an effort to get closer to statistical significance. We wanted to see if more data would result in IPM normalization (in other words, if the test results would settle back down to more even performance across the variations). However, the results of the second test remained the same. Note: the ad set(s) did not exit the learning phase and we did not follow Facebook's best practice.

  The results of this first test, while not statistically significant, were surprisingly enough to merit additional tests. So we tested on!
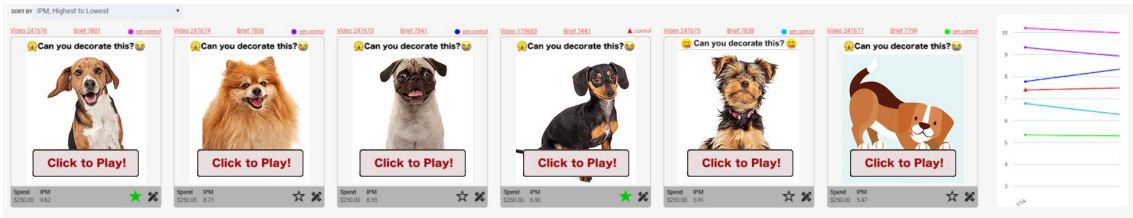
Second A/A test of video creative

For our second test we ran the six videos shown below. Four of them were controls with different headers; two of them were new concepts that were very similar to the control. Again, we didn't run the hotdog dogs; they've been inserted to protect the advertiser's identity and to offer you cuteness!

The IPMs for all ads ranged between 7-11 - even the new ads that did not share a thumbnail with the control. IPMs for each ad in the far right of the image.
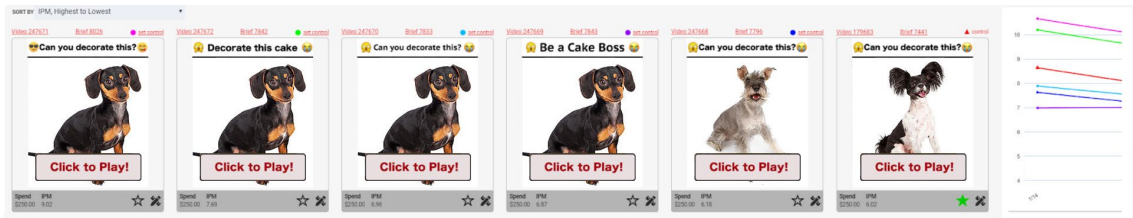
## Third A/A test of video creative

Next, we tested six videos: one control and five visually similar variations to the control but one very different to a human. IPMs ranged between 5-10. IPMs for each ad in the far right of the image.

Fourth A/A test of video creative

This was when we had our "ah ha!" moment. We tested six very different video concepts: the one control and five brand new ideas, all of which were visually very different from the control and did not share the same thumbnail.

The control's IPM was consistent in the 8-9 range, but the IPMs for the new visual concepts ranged between 0-2. IPMs for each ad in the far right of the image.
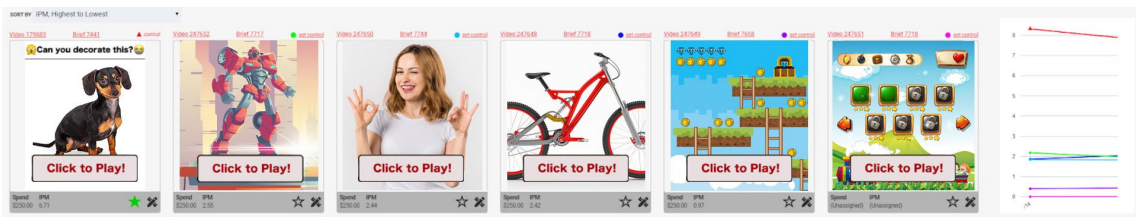
Here are our impressions from the above tests

- Facebook's split-tests maintains creative history for the control video. This gives the control a advantage with our non-statistically relevant, non-standard best practice of IPM testing.

- We are unclear if Facebook can group variations with a similar look and feel to the control. If it can, similar-looking ads could also start with a higher IPM based on influence from the control -- or perhaps similar thumbnails influence non-statistically relevant IPM.

- Creative concepts that are visually very different from the control appear to not share a creative history. IPMs for these variations are independent of the control.

- It appears that new, "out of the box" visual concepts vs the control may require more impressions to quantify their performance.

- Our IPM testing methodology appears to be valid, if we do NOT usea control video as the benchmark for winning.

IMP Testing Summary

Here are the line graphs from the second, third, and fourth tests.
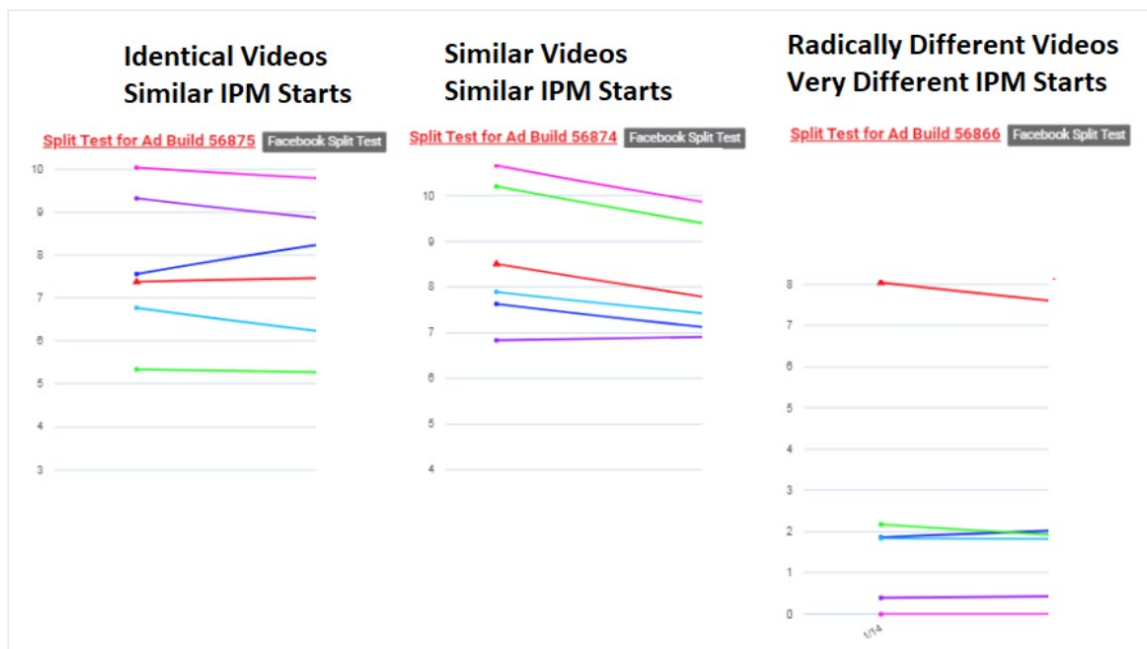
And here's what we think they mean:

Creative Testing 2.0 Recommendations

Given the above results, those of us testing using IPM have an opportunity to re-test IPM winners that exclude the control video to determine if we've been killing potential winners. As such, we recommend the following three phase testing plan.

Our 3-Step Creative Testing Process

1 IPM TEST
- New campaign for testing
- NO control video
- 1 ad per ad set
- FB split test, 3-6 creatives
- Drive 100 installs
- Either operating systems
- Newsfeed or Fan
- Winner(s) move to #2 for IAP titles

2 INITIAL ROAS
- New campaign AEO/VO
- NO control video
- 1 ad set, multiple ads
- Budget $500/day 7 days
- Either operating systems
- All placements
- 2-3 top audiences
- Winner(s) move to #3

3 ROAS SCALE
- Strong CBO Campaign (AEO / VO)
- NO control video
- New ad set
- Exit the learning phase
- Winner(s) back to Client for UA teams to scale

Phase 1: IPM Test

- No control video
- Create a new split test campaign using 3~6 new creatives (no control).
    – Setup campaign structure for basic App Install (No event optimizationor value optimization)
    – Spend an equal amount on each creative. Ex: One ad per ad set.
    – Budget for at least 100 installs per creative
        ° $200~$400 spend per ad is recommended (based on a CPI of $2-$4) if T1 English-speaking country
        ° $20~$40 spend per ad/adset testing in India (based on $0.20-$0.40 CPI)

- US Phase 1 testing.
    – 10-15% LAL with a seed audience similar to past 90-day installers, or past 90 day payers.
- Non-US Phase 1 testing.
    – Use broad targeting & English speakers onlyoIf not available in India, try other English-speaking countries with lower CPMs than U.S. and similar results. Ex: ZA, CA, IE, AU, PH, etc.

- Use the OS (iOS or Android) you intend to scale in production
- Use one body text
- Headline is optional
- FB Newsfeed or Facebook Audience Networking placement only (not both and not auto placements)
- Be sure winner has 100+ installs (50 installs acceptable in high CPI scenarios)
  - 100 installs: 70% confidence with 5% margin of error
  - 160 installs: 80% confidence with 5% margin of error
  - 270 installs: 90% confidence with 5% margin of error
- IAP Titles: kill losers, top 1~3 winners go to phase 2
- IAA Titles: kill losers, allow top 1~3 "possible winners" to exit the learning phase and then put into "the Control's" campaign

Which Creatives Move From Phase 1 > Phase 2?
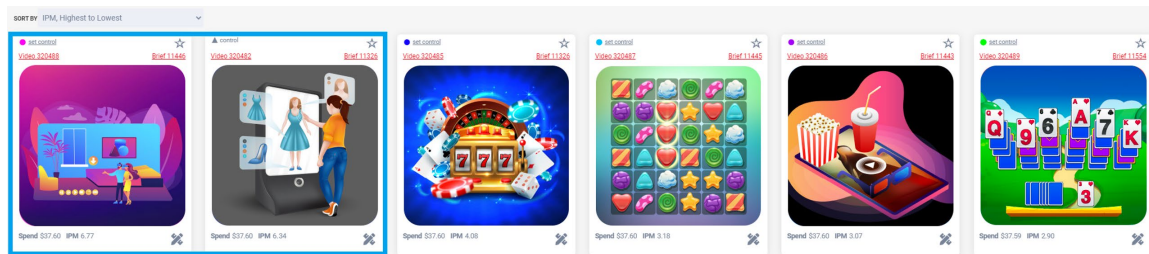
How to Pick A Phase 1 IPM Winner

- IPMs may range broadly or be clumped together
- Goal: kill obvious losers and test remaining ads in phase 2
- Ads (blue) have IPMs 6.77 & 6.34, move to phase 2
- If all ads are very close (e.g. within 5%), increase budget
- IAA (in app ads titles) you may need more LTV data before scaling

Phase 2: Initial ROAS

- No control video
- Create a new campaign with AEO or VO optimization
- Place all creatives into a single adset (Multi Ads Per Adset)
- Use IPM winner(s) from Phase 1 (you can combine winners from multiple Phase 1 tests into a single Phase 2 test)
- OS – Android or iOS. 5-10% LALs from top seeds (purchases, frequent users + purchase) + Auto Placements
- Testing can be done at a lower cost if you wish to run this campaign in other countries where ROAS is similar or higher but CPMs are much lower compared to US - ie. South Africa, Ireland, Canada, etc.

• Lifetime budget $3,500-$4,900 or daily budgets of $500-$750 over the course of 4-6 days (depending on your $/purchase).

WARNING! skipping this step is highly likely to result in one of the following scenarios:
- Challenger immediately kills the champion / control but hasn't achieved enough statistical relevance or exited the learning phase and therefore the sustained ROAS/KPI may not be sustained.
- Champion / control video has a lot more statistical history and relevance and most likely has exited the learning phases and may immediately kill the challenger before it has a chance to get enough data to properly fight for ROAS.

Phase 3: ROAS Scale

• No control video
• Use strong CBO campaign
• Choose winner(s) from Phase 2 with good/decent ROAS
- You've proven the ad has great IPM and "can monetize"
- To win this phase, it must hit KPIs (D7 ROAS, etc.)
• Create copy of an existing ad set
- Delete old ads and replace with your Phase 2 winner(s)
- Allows new ads to spend in a competitive environment
• Then, create new ad set, roll out towards target audiences with solid ROAS /KPIs
• CBO controls budgets between ad sets with control creatives and ad sets with new creative winners.
- Intervene with adset min/max spend control only if new creatives don't receive spend from CBO.
• Require challenger to exit the learning phase before moving to challenge the control "Gladiator" video
• Once the challenger has exited the learning phase, allow CBO change budget distribution between challenger and champion

Note: We're continuously testing our assumptions and discussing testing procedureswith large Facebook advertisers.

We look forward to hearing how you're testing and to sharing more of what we uncover soon.

FREE CREATIVE TESTING

- For the first time, anyone who becomes a client will have access to FREE Facebook Creative Testing.

- You can remove the burden of A/B testing creative from your internal UA team.

- You cover the media fees, and Consumer Acquisition's world-class UA team will manage the media buying of creative split tests for FREE (0% of spend) using our proven 3-step methodology detailed above.

- Contact: Sales@ConsumerAcquisition.com if you would like to take advantage of this offer.

FACEBOOK CREATIVE TESTING:

WHY DOES THE CONTROL ALWAYS WIN?

CONSUMER ACQUISITION

FREE CREATIVE INSPIRATION

- Want to know which of your competitors' video creative really drive their best performance?

- View +1,000,000 video ads from competitive apps and see which creatives work.

- To get full access to all FREE Creative Inspiration, please register for a free AdRules account or contact Sales@ConsumerAcquisition.com for more information.
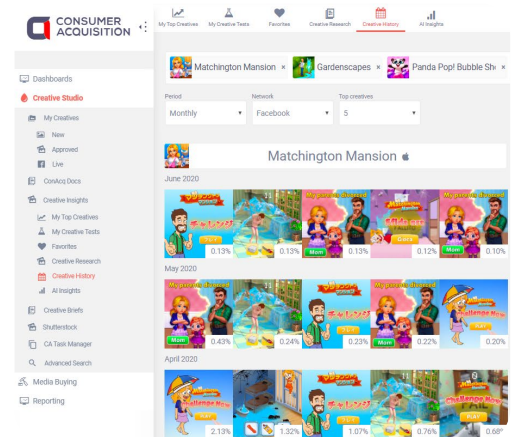
FREE MOBILE APP INDUSTRY BENCHMARKS

- Ever wonder how your mobile game or app KPIs perform vs industry benchmarks?

- Check out our "Mobile App Industry Benchmarks" dashboard and it is 100% FREE.

- See competitive KPIs like CTR, CPM, CPC, CPI, IPM, Conv%, country breakdowns, and more.

- To get full access to FREE industry benchmarks, please register for a free AdRules account or contact Sales@ConsumerAcquisition.com for more information.

Mobile App Industry Benchmarks