DDMA

White paper

Experimentation on websites and other online environments

The ultimate guide to the comprehensive design and analysis of experiments, and creating an experiment-driven corporate culture

Foreword

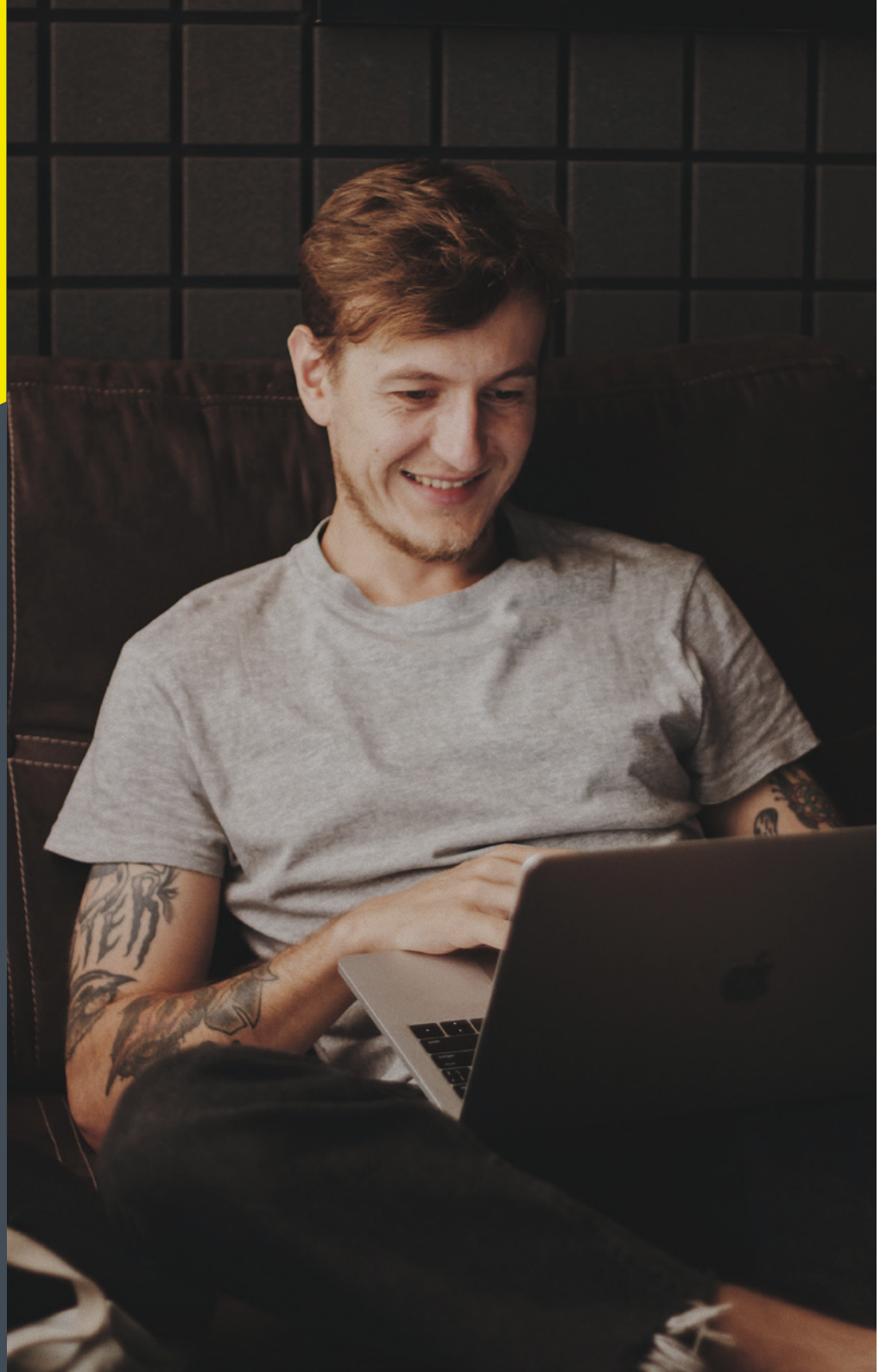
In a world where technology is constantly evolving and the market is becoming increasingly competitive, it is crucial that organisations adapt to changing circumstances. One of the most effective ways to discover what is necessary in order to do adapt is through experimentation. By conducting structured experiments on websites and online platforms, hypotheses can be tested and valuable insights gained regarding the behaviour of your users and therefore the impact of changes in terms of processes, products or services.

It is therefore our greatest pleasure to introduce this white paper by the DDMA Experimentation & Optimisation Committee* (formerly DDMA CRO Committee), based entirely on our belief that experimentation is a powerful tool for improving business performance and achieving objectives, as experimentation offers opportunities to make fact-based decisions and to prevent mistakes. The white paper describes in clear steps how best to conduct an experiment.

However, it is not enough to merely know how to conduct an experiment. Today, the importance of experimentation extends to all facets of an organisation. Whether it's marketing strategies, product development, website design, or operational processes, experimentation allows you to make data- based decisions and to constantly evolve. That is why, in this white paper, we also discuss the importance of iteration, and creating an experiment-driven culture within the entire organisation.
I sincerely hope that this white paper will inspire you to embrace experimentation as a strategy to transform and grow your business

Frank Overbeek,
DDMA Experimentation & Optimisation Committee Chairman

*The DDMA Experimentation & Optimisation Committee focuses on proving the value of data-driven experimentation and testing as a means for well-rounded decision-making within organisations. The committee offers organisations knowledge-sharing, guidelines, and models for inspiration and support in the growth concerning the maturity of their experiment team and culture."



Inhoudsopgave

Introduction

Experimenting and testing online environments is vitally important for any organisation. However, not everyone is aware of this yet, or knows how to approach it in a strategic manner. In addition, people often experiment from silos, which leads to experiments with (too) little support from within the organisation.

Enough reason for us, the DDMA Experimentation & Optimisation Committee, therefore to dedicate a white paper to the ways of approaching an experiment (series) strategically, by utilising a working method that benefits the entire organisation.

What is classified as an experiment?
In this white paper we use the following definition of an experiment: an experiment is the substantiated observation and testing of a predefined hypothesis.

Why experiment on your website or platform?

Before we go deeper into the subject matter, it is important to know what exactly the direct and indirect value is of experimenting on a website or platform, and creating an experiment-driven culture. What exactly can you derive from an experimentation strategy?

Well, a lot of job satisfaction for one thing. Academics from other experimentation disciplines typically envy website owners who can conduct online experiments. Due to the often enormous sample sizes, the segmented sample size, and the low incidence of research bias, online quantitative testing is a proper party. Website visitors are also – once they give permission – ideal respondents, because they immediately provide website owners with the feedback needed to test and ultimately implement changes to their platform. This results in a win-win situation, whereby on the one hand website visitors benefit from a better, smoother and more customized website experience, and website owners on the other hand benefit from better results.

.

Roughly speaking, the value of experiments can be expressed through the following components:

1. **Direct financial impact**
   - Result from winning experiments
   - Avoiding risks due to a "negative" outcome
   - Increased retention

2. **Indirect impact**
   - Increased customer satisfaction
   - Customer Effort Score improvement, Net Promoter Score, Word of Mouth Marketing, etc.

3. **Product owners within your organisation will gain a fuller and deeper knowledge regarding end users, and their behaviour and attitude**

4. **In general, more knowledge within the organisation concerning the value and outcomes of experiments, and the skills needed to optimise an experiment**

Valuable experiments provide substantiation for important business decisions

Valuable experiments are often the deciding factor in an urgent or important business matter. The result therefore extends beyond merely monetary KPIs. For example:

- **Logistical challenges**: such as with the deliveries of products, how do you direct visitors to desirable alternatives on the web page, for example, if normal delivery is not possible? How are you looking to solve that as an organisation?
- **Internal deadlocks**: Sometimes departments can bicker internally about rights and wrongs. For example, if the legal team wants to take a different approach to marketing over purchase terms and conditions.

In these kinds of situations, an experiment (series) can provide just the evidence needed to lend direction in unclear situations and set things in motion again. In addition, there are also valuable knock-on effects of experimenting on your website. For instance, experiment results generate well-founded discussions about platform changes. This saves discussion time, and also contributes towards the adoption of data-driven decision-making. It's a worthwhile side effect, especially for organisations that still make multiple gut feeling-based decisions.

What if there is no experimentation?

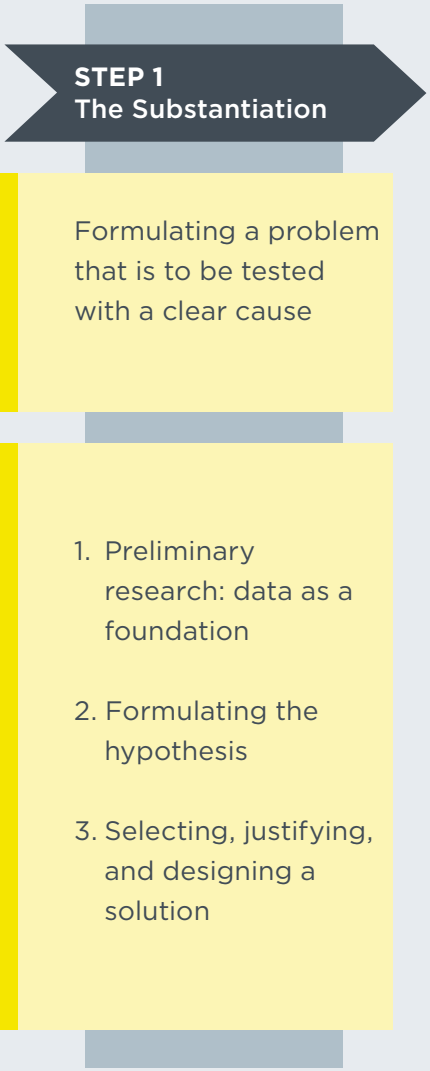
Making changes to your platform can have a major impact on business objectives. If you validate those changes in a thorough manner, you will know the initial cause of those consequences. It's crucial to do, especially because on average, a large part of all changes have a negative impact on the results. Validating the changes will prevent losses and also teach you why those changes fail to work.
In other words, if you fail to validate any - or merely a small part - of your platform changes with experiments, it is important to do it as soon as possible; this is because it's the only way to:
1. Learn more about your customers
2. Create more valuable experiences for your end users
3. Limit your potential losses

Use the tools from this white paper for your own business operations

It should be clear that experimentation is – or at least should be – a crucial part of any organisation. In this white paper, we focus primarily on how to achieve a successful experiment. We will take you through how to carry out an experiment in 4 steps, from start to finish: Substantiation (1), setup (2), analysis (3), sharing and support (4). Finally, through two in-depth chapters, we will discuss how you can make experimentation a structural part of your business practice.

Steps of the experimentation process

	STEP 1 The Substantiation	STEP 2 The Setup	STEP 3 The Analysis	STEP 4 Share the Experiment
OBJECTIVE	Formulating a problem that is to be tested with a clear cause	Ensuring a sound technical setup and usable results	Ensuring a robust experiment with valid conclusions	Building support for the results and implementation of solutions
ACTIONS AND FOCAL POINTS	1. Preliminary research: data as a foundation 2. Formulating the hypothesis 3. Selecting, justifying, and designing a solution	1. Determining the sample size Using: - Baseline conversion - Confidence interval - Power - Minimal Detectable Effect (MDE) 2. Determining the experiment duration	1. Hygiene factors analysis - Error check - Preventing Sample Ratio Mismatch - Hygiene check 2. KPI analysis	1. Transparency and collaboration 2. Communication is key 3. Thorough documentation



STEP 1

The substantiation of an experiment

The quality of an experiment is dependent on two crucial aspects: solid substantiation and a thorough design. Three things are essential in the substantiation of the experiment: the quality of the preliminary research, the hypothesis and the relationship between the solution found in the experiment, and the preliminary research. We will discuss these 3 issues in detail in this chapter.

1.1 Preliminary research: Data is the foundation

A good experiment is characterized by thorough preliminary research with a well-considered use of data. It is key to combine the different data sources you use in the right way – this is also referred to as data *triangulation*. Previously, data sources consisted mainly of web analytics reports. Nowadays this is no longer sufficient, and you need to mix and match multiple sources. Think of combining meta-analyses of previous experiments with data from surveys, feedback e-mails, heat maps, customer contact centres and other quantitative data sources. This way, you create the broad foundation needed to build a good hypothesis.

However, triangulation only occurs when data sources complement each other – rather than just confirm each other. You do not then seek out sources that merely show the same thing. Instead, you are looking to match data sources with one another like puzzle pieces. For example: not only data sources that show that users exit in the check-out, but also data sources that show why they decide to exit there. You need to look further, and connect the insights from different data sources in such a way that the conclusion is more than the sum of its parts. In addition, you can also opt to include qualitative data in your experiment, which lends context to the figures, and renders your substantiation even more solid.

To give an example: data from an organisation's complaint register shows that there is uncertainty surrounding delivery terms and conditions. Heat maps show that delivery terms and conditions are rarely displayed. The web analytics reports show that there is a high exit rate on the page where visitors should be able to find these delivery values. In this way, you combine data from different sources to arrive at the same conclusion through various pathways.

The danger of confirmation bias

It is extremely important that while interpreting the data in relation to determining the conclusion for the hypothesis, you remain honest with yourself and the company. There is always a chance of falling into confirmation bias, whereby you subconsciously look for data that coincides with intrinsic, desirable expectations. Without realizing it, you might fail to see or ignore certain patterns to ensure that the outcomes align with your own points of view. The *degree of subjectivity* of the assessment then dominates the proceedings, which can be detrimental to the outcome. So, take care to adhere to the experiment standards and best practices within the field as much as possible.

1.2 The components and preconditions for a good hypothesis

One way to avoid confirmation bias is to create a pure hypothesis. A perfect hypothesis contains a number of components:

- Target audience
- Modification/adaptation
- Expected change in behaviour
- The underlying reason or a consequence of the expected change in behaviour (optional)

The preliminary research continues in a natural way in a hypothesis that is to be validated, and the final experiment variant(s) actually test what is proposed by the hypothesis. An experiment that shows a tidy profit, but that cannot be traced back to the hypothesis, may result in a desired increase in a chosen KPI, but in terms of knowledge and learning capacity, this does not contribute in a desirable way. A seamless connection between hypothesis and experiment variant is therefore crucial for a successful experiment, but can be challenging.

Two examples of a good hypothesis

The highlighted sections from the hypotheses below correspond to the colors in the legend on the left side of this page.

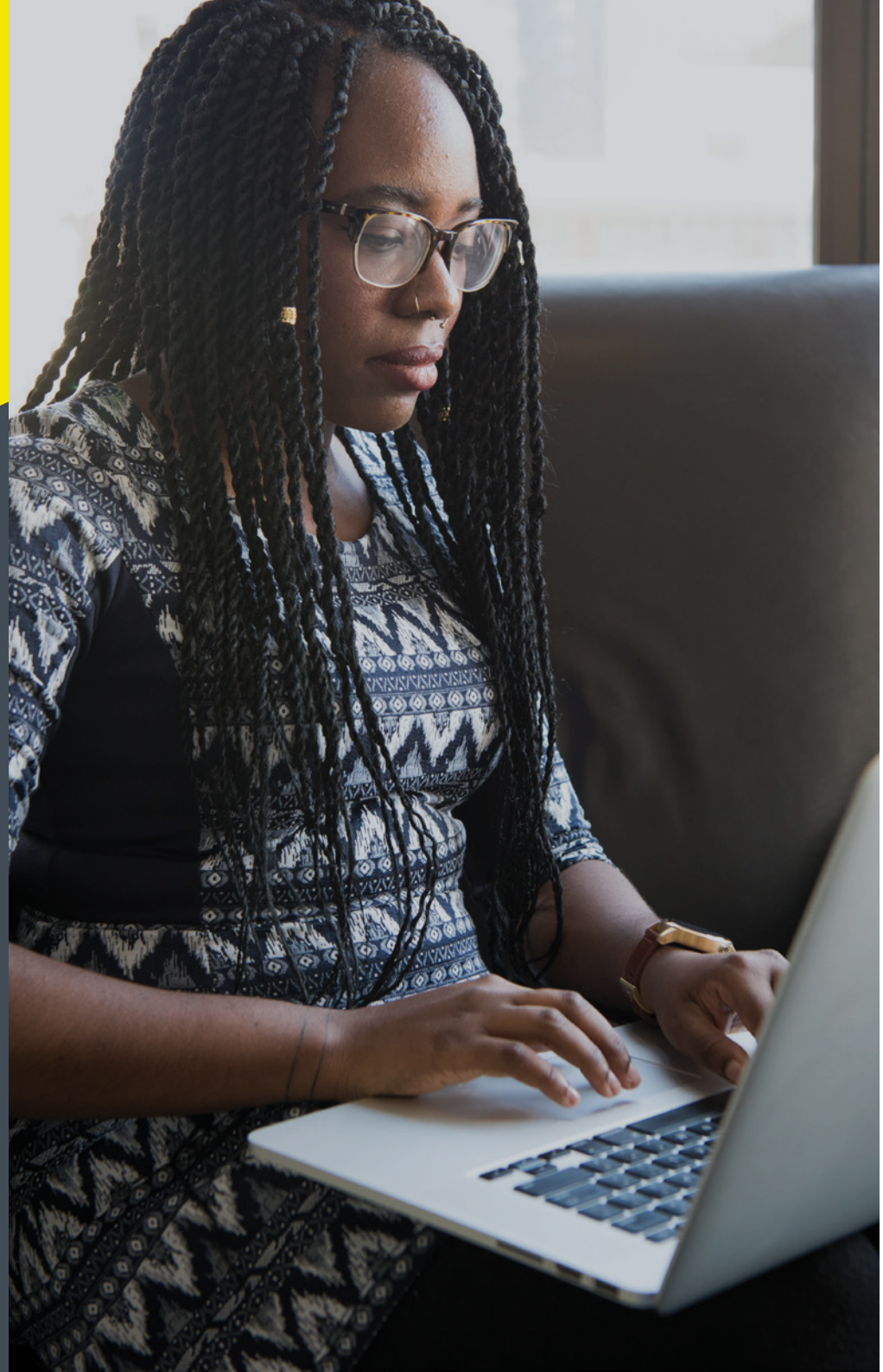
"By removing the barriers surrounding delivery in the check-out, the ability of the new customers is increased and therefore the conversion rate increases".

"If we reduce the cognitive effort required when choosing the delivery date, then the last step of the check-out can be completed more smoothly. As a result, we observe lower exit rates and higher conversions."

The second example is the hypothesis of a winning experiment by the Beter Bed company. In the end, Beter Bed managed to win the DDMA Dutch CRO Award in 2022 in the Conversion category with their experiment series. Curious about the case? Check out the case here.

A few preconditions for a good hypothesis

- A hypothesis should define the necessary parts as clearly as possible;
- A hypothesis is preferably written 'in one direction', and clarifies what you are going to validate. In other words, not: the expectation is that the conversion rate will change, but instead: the expectation is that the conversion rate will increase;
- A hypothesis preferably does not yet contain the solution, since it's possible to confirm a hypothesis in multiple ways and with multiple variants. Hypotheses with solutions such as *'by*



making the button text "buy now"...', are therefore not preferable. You can keep it more open by formulating solutions more broadly: '*by adjusting the copy...*';
- A good hypothesis revolves around improving one primary KPI, so you don't have to formulate multiple expected effects within one hypothesis. In step 4 (page 21) we take a deeper look into KPIs.

1.3 Select and substantiate the solution

Once you have a hypothesis, you can look for the visual reflection of the solution you are looking to test in the experiment. Here too, it is crucial to thoroughly substantiate why the chosen solution is the most suitable, and why you opted for that one instead of another. The combined data sources that you have previously used in the preliminary research and the preparation of the hypothesis can be put to good use in this regard. Use that data to show that there is a direct relationship between the problem and the proposed visual solution.

When devising and visually designing a suitable solution, it is of great importance to take into account the various interests of any stakeholders. Drawing up a stakeholder map can help to gain insight into the different goals, interests and possible concerns that each stakeholder has. By incorporating this information into the experiment design, it is easier to generate more support for the chosen solution in the event of a positive result. We will talk about stakeholder management in more detail in the in-depth topic 1 (page 24).

DDMA

STEP 2
The Setup

OBJECTIVE

Ensuring a sound technical setup and usable results

ACTIONS AND FOCAL POINTS

1. Determining the sample size

   Using:
   - Baseline conversion
   - Confidence interval
   - Power
   - Minimal Detectable Effect (MDE)

2. Determining the experiment duration

STAP 2

The setup of an experiment

You now have the hypothesis and the solution in mind that you want to test in your experiment. Now it's time for the setup. A good experiment stands or falls by its technical setup. If the setup is right, the results and conclusions will be useful, and you can implement the solution to make an impact within your organisation. But how do you set up a good experiment? We will explain about that in this next chapter.

When setting up an experiment, you need to think about the following:

1. Determine the required sample size of the experiment, determine the following matters for this purpose:
   - The baseline conversion
   - The confidence interval
   - The power
   - The Minimal Detectable Effect

2. Determine the required duration of the experiment
   Consider the following factors:
   - Expose control and test variants to the same target groups
   - Pay attention to consistent target group sizes in follow-up tests

2.1 Determine the required sample size

For a successful experiment, the sample within which an experiment falls must be large enough to observe a significant effect. To determine the sample size, you need to go through the following steps:

1. **Determine the baseline conversion**
   Start by establishing baseline conversion. That's the current conversion rate you want to improve upon. This can be based on historical data or market research. A higher baseline conversion leads to a smaller sample size requirement because the effect is easier to detect.

   Contrary to the other three subsequent components for determining sample size, the baseline conversion rate is derived from data. The baseline conversion rate is therefore fixed, and you cannot change it in the setup of an experiment.

2. **Determine the reliable interval**
   To ensure that the result of the experiment is based as little as possible on chance, it is important to determine a confidence interval. The narrower the confidence interval, the more precise the results, and the larger the sample size. A significance rate of 95% or 99% is common. A significance of 95% indicates that in 95 out of 100 cases, the results correspond to the predetermined hypothesis, i.e., the probability of a different result than expected is less than or equal to 5% ($p \leq 0.05$). In general, for a thorough experiment, the higher the signification, the better.

3. **Determine the power**
   The power is the ability of the experiment to detect an effect when it is actually present. The power is usually expressed as a percentage, often with a target value of 80% or higher. A power of 80% means that the experiment will detect an effect in 80% of cases if that effect is genuinely present. In other words, the experiment has a good chance of producing correct positive results. It's important to have sufficiently high power to ensure that your experiment is able to accurately detect effects. Power that is too low can result in the missing of key insights or drawing of wrong conclusions. When calculating the required sample size, use the power in combination with other statistical parameters, such as the expected effect size and confidence level, to determine the minimum sample size needed to achieve the desired power level.

4. **Determine the minimal detectable effect (MDE)**
   That is, the smallest effect you want to be able to observe and explain in a statistically significant manner. We refer to this as the minimal detectable effect (MDE). If the MDE is large-scale, then a limited number of visitors is needed to determine a winner of an experiment and vice versa; the smaller the MDE, the larger the sample size must be to achieve statistical significance. For example, a change to a website button text can result in a change/effect of 5%, while a change with a completely new website functionality can achieve an effect of 20% for example. However, this can also be completely the other way around (small changes with large differences). It is therefore important to map the MDE in relation to the intended change, and to include it in the calculation of your sample size.

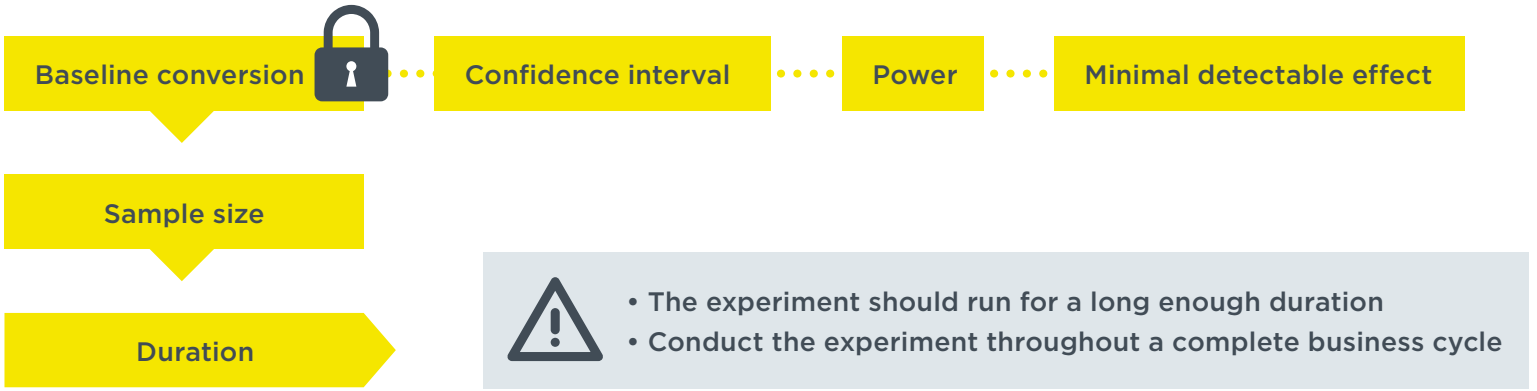
There is no exact science in determining an MDE, and therefore customisation (based on the aforementioned changes and historical experiment results) is required. The common starting point in the market is a percentage between 5 and 10 percent.

2.2 Set the duration

If the size of the required sample is clear, you can make a realistic estimate of how long an experiment will take to reach the required sample size. In other words, the duration. Keep in mind the following:

1. **An experiment must run for long enough**: In order to collect enough data points, an experiment must run for long enough, but not too long however, because external factors can then influence the results. Think, for example, of visitors who delete their cookies and are therefore marked as new visitors, or changes in behaviour due to the changing of seasons.

2. **Experiments must be run in full business cycle units**: this way, you prevent a contaminated test as the result of peak moments. For example, run a test in full weeks of 7 days (instead of 5), unless the business cycle deviates from this – for example, in monthly lotteries. There is also often seasonality at play at day, week or month levels. For example, there are more purchases at weekends, or when salaries have just been paid.



There are various tools available that can calculate the duration for experiments. A 2-4 week period is usually an acceptable duration for an experiment. If it turns out that an experiment requires a longer duration, then pay attention to whether the experiment setup is receiving enough visitors, and what period of time is needed to arrive at the desired sample size.

2.3 Limit the number of element changes per follow-up experiment

When setting up experiment variant(s), it is important to think about the number of elements that you change per follow-up experiment. If there are too many per follow-up experiment, it can become very difficult to demonstrate why, how and which change has resulted in an ultimately significant result. In addition, changes with opposite effects may result in nothing being visible at experiment level. Therefore, it is recommended to set up one element change per experiment with a clearly measurable KPI. If you keep multiple element changes, the sample size increases.

2.4 Expose the control group and the test variant to the same target groups

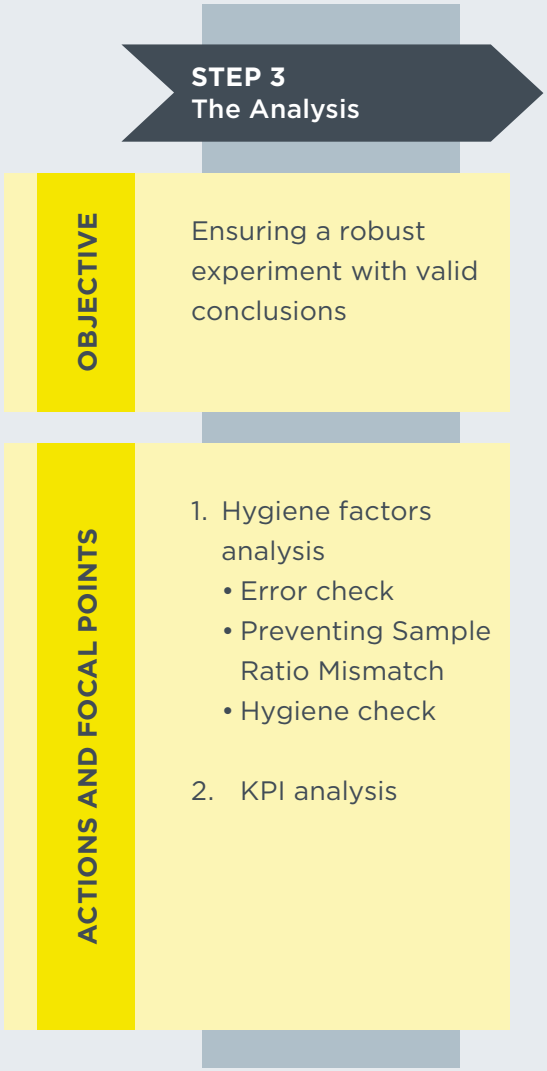
Both the control group and the variant of the experiment must be exposed to similar (randomly allocated) target groups. An experiment in which different RFM* segments are shown a different experiment variant is very difficult to analyze in a pure manner – unless you also put a control group within those segments. When segmenting, it is also important to include the impact on the sample size (reduced by the segmentation) in any calculations.

2.5 Maintain the same sample size in follow-up tests

Both the default variant and the follow-up variant(s) of an experiment must be performed on the basis of target group samples of equal size. For example, a 10/90 distribution could lead to problems during the analysis, and is not preferable in the experiment setup. In case of risk avoidance, you can opt for a 10/10/80 split, where you allocate the default and variant both 10% traffic, and keep the remaining 80% of the traffic out of the experiment.

*RFM stands for recency, frequency and monetary value respectively. Based on this, visitors/buyers are often divided into segments on the basis of behaviour. For example: how recently did you make a purchase, how often, and for how much money?

DDMA



STEP 3

The analysis of an experiment

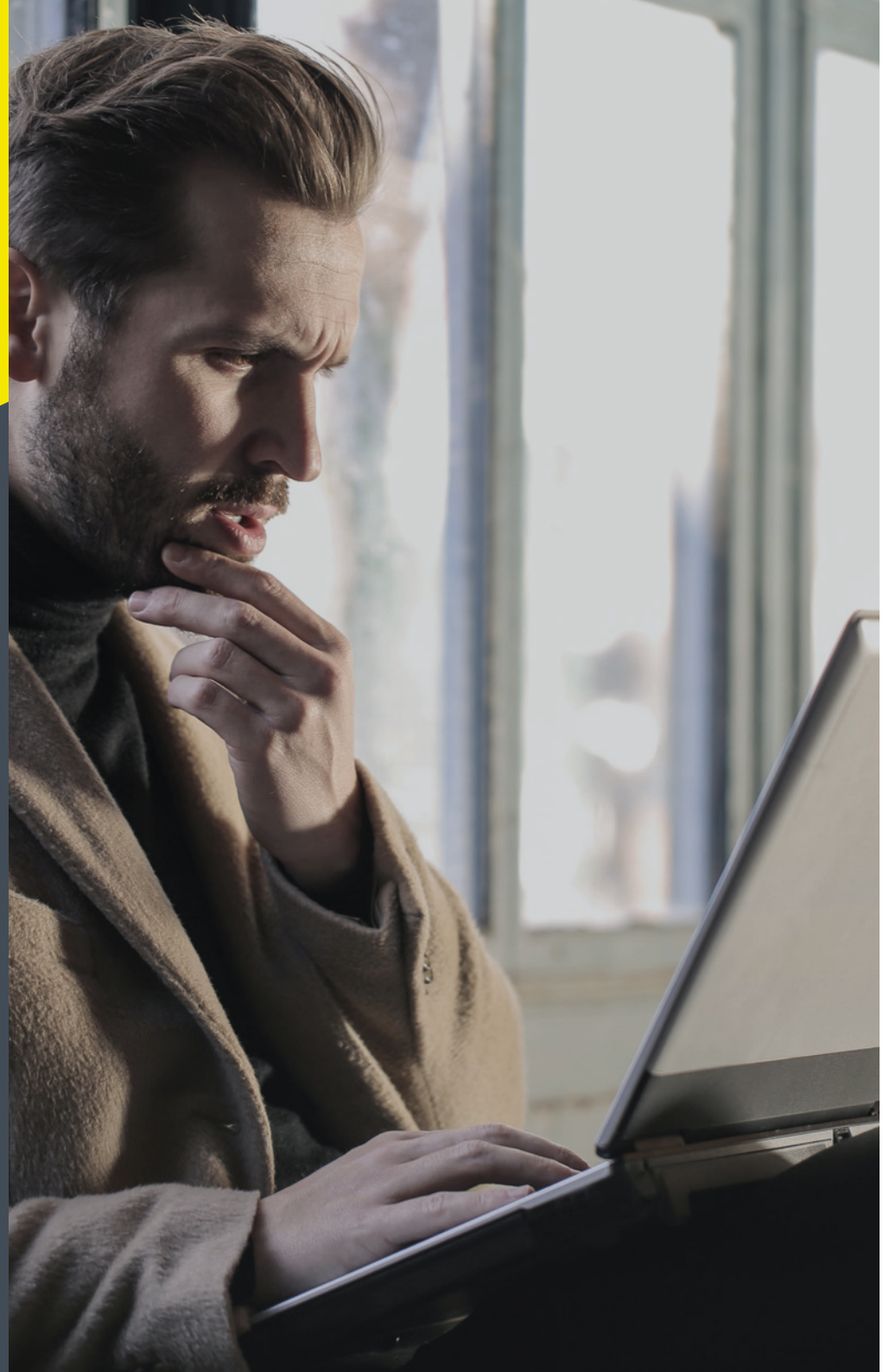
It is crucial that the analysis of an experiment is done carefully, otherwise, you run the risk that results are not valid, and therefore not usable. Or that you implement an intended solution based on an invalid result, while it has a negative impact on what you ultimately want to achieve.

To ensure valid and useful results, two points are important:

- The hygiene factor analysis
- The KPI analysis

3.1 Hygiene factor analysis

A hygiene factor analysis helps identify potential obstacles that could negatively affect conversions, such as poor navigation, slow load times, and errors in website functionality. So, it's actually an indicator for the health of your experiment.
By identifying and solving these obstacles, you can be sure that the numbers that roll out of the experiment are valid. A hygiene factor analysis comprises the following components:



1. **An error check**
   Data that comes in from an experiment must be clean. Therefore, pay close attention to errors and carry out checks. For example:
   - **Filter the IP addresses of employees from the data**; This prevents experiments from becoming contaminated with data from your own employees;
   - **Do checks for any bugs in the control or test variant**: For example, if it turns out that a certain button was not clickable, or forms do not run smoothly, this can affect the results;
   - **Properly map outliers in the data**: You might reach positive significance very quickly with a limited number of conversions, but don't stop the experiment! This can be due to (external) influences that can even be negative in the long run. Therefore, monitor conversion trends and only draw conclusions when they are evenly distributed over the entire duration.

2. **Preventing/Resolving a Sample Ratio Mismatch**
   If there is a Sample Ratio Mismatch (SRM), then the size of your sample (distribution) will differ significantly from the size/distribution you initially envisaged. This can lead to biased results, as the uneven distribution of users can affect the performance of a variant. It is essential to ensure that the samples in the experiment groups are representative and balanced in order to draw reliable conclusions.

   If the analysis indicates that there is a sample ratio mismatch (SRM), this means that a certain group of users may be overrepresented in the experiment. This can happen, for

example, because more returning visitors participated, or because certain browsers or device types (such as mobile users) were underrepresented in a certain variant.

The presence of SRM can make the detection of an effect in the experiment easier or more difficult, depending on the associated degree. To understand the influence of SRM on the experiment results, additional research is needed. In short, if you observe an SRM, it is necessary to investigate why this happens, and determine whether the results of the experiment are still valid after any possible causes of SRM have been addressed. This helps in obtaining a more reliable picture of the effects of the experiment.

3. **A hygiene check at channel, device and segment level**
   A hygiene check is essential to ensure that the results are not affected by unforeseen factors, and to facilitate an accurate evaluation of the experiments. This can be done at the channel/traffic, device, and segment level.

   - **Channel and traffic**
     Ensure an equal distribution of marketing channels between the default and variant(s) to prevent the effectiveness of a specific channel from affecting the results. If the variant of your website achieves much better results with e-mail traffic, while the standard version performs better with display ads, this does not necessarily mean that the variant itself is more effective. It really only shows that e-mail as a channel is more effective than display. It's important that you make this distinction in reporting by showing separate data for each channel so you can see how each channel is performing. This allows you to understand whether the difference in conversions is due to the variant or the channels used.

   - **Apparatus**
     Analyse the results by device type, including different browsers per device type, to understand possible differences between devices and browsers. It may be that a found gain on mobile fully compensates for losses on

Take an A/A test and a retest to solve an SRM

In an A/A test, you perform an extra experiment whereby both variants are identical to each other. The purpose of this is to find out the structural cause of the SRM and solve it, usually by making adjustments to the setup of the tooling of the experiment. By observing whether the SRM is still present in the identical variants, you can better understand what causes it and address it in a targeted manner.

If it is clear that an experiment is contaminated with SRM, you can choose to perform a retest. Here you compare the variant again with the default option to confirm whether the result found is reliable. If there are still significant differences between the variants, this indicates low reliability of the experiment. In that case, it is necessary to conduct further research into the cause of the problem.

desktop, then it is desirable to find out where those losses on desktop come from in order to get (even) more out of segmented follow-up experiments. After all, experimentation remains an iterative process. You can read more about this in the in-depth topic 1 (page 24)

- **Segments**
  Always include those segments specified in the hypothesis in the analysis. Consider looking at other segments not specified in the hypothesis to assess whether the experiment may have undesirable effects on specific groups. It is possible that an experiment shows a nice uplift, but that this is solely generated by existing customers. You may then be overlooking the fact that you are driving away potential new customers, despite the fact that the experiment is producing a positive result.

Performing these checks helps to get a more complete picture of the results and possible impact of the experiment.

3.2 The KPI analysis

Each experiment must ultimately be assessed on the pre-formulated KPIs. It is important that you genuinely determine the primary KPI in advance and stick to it in the analysis (see also chapter 2). If you evaluate an adjustment on the homepage based on each available KPI within a transaction, you will make the analysis too complex and possibly less pure. A hypothesis should therefore be written with one main KPI as a focus, and that should be utilised for the analysis on that KPI.

3.3 Which A/B test method do you select?

There are two common statistical methods for analysing and interpreting the A/B test results, namely the Frequentist and Bayesian methods.

1. **Frequentist method**: This method is a way to draw statistical conclusions based on collected data. It's all about looking at the number of times an action takes place. Using this method, you draw up a zero hypothesis – a kind of standard assumption – and an alternative hypothesis that differs from that. Following that, you collect data and look at how likely it is for that data to fit the zero hypothesis. You determine whether the results are statistically significant, which means they are unlikely if the zero hypothesis is true. Using this as a basis, you then draw conclusions about which variant of an experiment is a statistically better performer.
2. **The Bayesian method takes into account** within the calculation that there was already an existing division on which control and variant perform prior to the experiment. The method looks at the (1% to 100%) probability percentage that your hypothesis will display a successful performance between the experiment variants. The higher the percentage, the more certain you can be of a successful result. Ultimately, you decide the acceptable percentage for trusting the assumption. In other words, you draw a conclusion about the probability that one of the experiment variants will render a better performance than the other.

Selecting the right method depends on several factors, including the specific experiment, the available data, and your own preferences. Both methods have their own advantages and areas of application, and it is important to select the right method based on the context and purpose of the experiment.

STEP 4

Share the experiment within your organisation and generate support

It worked. Your experiment has been carefully set up and analysed – and has generated results! You are keen to implement the envisaged solution within your organisation. But before you are able to, you might have to deal with interested stakeholders, whom you will have to convince first. Three factors are important in this regard: transparency and cooperation (1), good communication (2), and thorough documentation (3). Is everyone on board? If so, you can convert your experiment results into action points without hurdles, to implement the intended solution.

4.1 The importance of transparency and cooperation

All the experiments you perform never all lead to the intended result. Results are not always significant. It is key to realize that this is very normal, and an experiment failing is all part of the process. An unsuccessful experiment can open doors to continue testing with new insights. The road to a successful experiment with multiple tests shows the steadfastness and passion for a certain way of thinking, which is exactly why it is so important for transparency to surround the experimentation process. The journey to an ultimately successful experiment, and the accompanying steps and results ultimately lead to appreciation within the company and other stakeholders.

In addition, cooperation is also an important key to success. Working together between different business departments to jointly arrive at insights and results during an experiment not only increases the quality of the research, but automatically ensures more visibility across the board. As a result, experimentation teams aren't working on a remote island, but are integrated throughout the entire organisation, making it easier to reach and win over the top layer. You can read more about that in the in-depth topic 2 (page 26).

The experimentation-case van Beter Bed is a great example of the importance of transparency and cooperation. This involved working closely with the supply chain department to improve the ordering and delivery experience, resulting in valuable recognition at top management level.

4.2 Communication is key

No matter how cliché it might sound, communication really is key. Consistently involving the entire organisation by clearly communicating experiments can help colleagues outside the experimentation teams feel engaged and responsible during the experiments. This can be achieved through two simple implementations in your working method:

1. Choose fixed times for sharing experiment results.
2. Gather up any ideas from other teams for (follow-up) experiments

Ensure that the decision-makers (managers, etc.) within the organisation are included during these communication moments - especially when it comes to experiments that have a genuine impact on business objectives and results, as this can have a positive effect on the awareness of the experiment-driven culture within the company in the medium and long term.

4.3 Provide thorough documentation

The systematic and consistent storage and clear categorization of the experiment results ensures that you can learn from previous experiences, that no duplicate experiments go live at the same time (instead of consciously retesting something), as well as increasing awareness and knowledge- sharing within the company. A thorough documentation of all steps is therefore crucial for the learning process, an effective experiment project, and a mature experimentation culture. You can read more about mature experimentation culture in the in-depth topic 2 (page 26).

A good experimentation specialist therefore looks beyond KPIs, and also at the bigger picture. An experimentation team must therefore work with measurement plans at all times, whereby business objectives are broken down into sub-goals with sub-KPIs. It is crucial that you always keep the interests of all stake-holders in mind. What valuable insights do they need that you can get from your experiment? A good start is to create a stakeholder map, with which you create an overview of the interests and power of stakeholders with regard to your experimentation programme.

4.4 Implementing experiment results

If an experiment variant clearly emerges as the winner, and you have convinced all interested stakeholders of it, you can proceed to the actual implementation of the solution in the online environment. If this doesn't happen, the test series has been done in vain. To do this, it is important to work closely with web development teams and any product owners/managers. Discuss together which results and solutions you will definitively implement, and which results need a follow-up experiment. You can also opt to make adjustments in iterations live, and to perform follow-up experiments as well.

In-depth topic 1: Strategic experimentation is an iterative process

A good experiment strategy never consists of solely one experiment. How do you devise such a strategy? We will discuss that in this chapter, where we elaborate on the value of experiments in business challenges, the power of a series of experiments surrounding a specific topic or theme, and how a continual focus on your problem can increase the chance of success. After all, experiments are not only valuable for testing tactical matters (bottom up); the same applies to strategic considerations (top down).

Q1.1 An experiment is never a one-off

We said earlier: an experiment is never a one-off. A valuable experiment often stems from the learnings of previous experiments. After all, experimentation is an iterative process. Step-by-step learning and acting in accordance in new experiments leads to limited mistakes. Ultimately, what you're aiming for is to arrive at an experiment that leads to a significant result after multiple incremental improvements.

Beter Bed launched a series of experiments in 2022 whereby they repeatedly tested their commencing date module. Prior to the last, winning variant of the experiment, 6 other variants were tested with a slightly different approach per variant. The result is therefore an iterative process of trying, learning, honing and trying some more.

In the end, Beter Bed won the DDMA Dutch CRO Award in the Conversion category. Curious about the details

Check out the case here.

V1.2 Experiment in sequences to rule out coincidence

The example of Beter Bed shows that you should always look at an issue from multiple angles. Just because you can't prove your hypothesis all at once, doesn't mean you should reject it right away. It offers an opportunity, which you as an organisation can take advantage of, by conducting new experiments. Just because you don't find a solution doesn't mean the problem doesn't exist.

But be vigilant, because one swallow doesn't make a summer; there is always a chance that a winning test is not a winner. In that case, the outcome is coincidental, and there is no winner. This is also known as 'false positive'. So it's a good idea to repeat winning tests (after a period of time) to be sure of your case. Experimenting in series therefore leads to valuable insights and a higher success rate. A positive side effect is that it gives you more focus as an experimentation team. It also provides an overview and the capacity for multi-disciplinary teams to tackle a major and important topic thoroughly.

Meta-analyses to learn more about topics

Experimentation in series does not always have to take place in a given period. Through a *meta-analysis*, you can draw conclusions about experiments that you have already done. For example:

- Run an analysis on all the experiments you've conducted on mobile.
- Run an analysis on experiments related to FAQs.

In order to be able to do this thoroughly, it is important that you record your experiments properly and consistently in your working method. It is advisable to tag at least the channel, device type, page type, KPI, result (positive significant/negative significant/no difference) and theoretical principles (such as Cialdini) per experiment. There are several good tools available for this purpose on the market, including Airtable, Jira, Effective Experiments and MetaDimensions. Often it is more than worth the investment, because it allows you to record results sustainably – clearly arranged in one location, and is easy to share with other teams.

In-depth topic 2: Creating an experiment-driven corporate culture

The maturity of your experiments depends on an optimal combination of several factors. This entails sufficient budget, time, knowledge, the right people and appropriate tooling and data. In summary, you can classify these factors under 4 aspects that are important for a mature experiment culture: (1) culture, (2) support & trust, (3) business objectives and (4) process.

V2.1 A culture of knowledge-sharing

Ongoing knowledge-sharing between teams takes place in only 13.5% of organisations. In these kinds of organisations, where there is a widespread enthusiasm for experimentation, experiments generate a flywheel effect, through which more and more colleagues contribute ideas and insights and ideally also experiment themselves. This cross-pollination is an ideal basis for an experiment-driven corporate culture.

Read more about the importance of knowledge sharing within a mature experiment programme in this article:
The current task assigned to the CRO specialist

V2.2 Support and trust

A crucial factor for success within an organisation is generating support. Even with unlimited resources, an organisation will not achieve success without the support of management. This requires the development of a culture within which the exchange of knowledge and data throughout the organisation leads to insights, understanding and the motivation to experiment.

In 54% of organisations, experimentation is supported by top management. This support is essential to ensure that experimentation is also applied to profound and bold changes. While you can still run an excellent experiment without this support, you run the risk of other stakeholders enforcing greater and unvalidated changes. A culture whereby top level management supports experimentation, and where results are shared with all relevant stakeholders within the organisation, often leads to new initiatives to further experiment along the successful pathway that has been embarked upon.

With this backing, you also have the required support and confidence to validate hypotheses of which - although substantiated with data, their outcome is still uncertain. This trust is also necessary to recognise the losing variants as learning points that have saved the organisation from detrimental choices.

V2.3 Business objectives

An experimentation programme generates the greatest possible impact by directly contributing to the realisation of business objectives. By ensuring that all steps within an experiment, from the initial data exploration to the final implementation of the winner, are in line with business objectives, you experiment with maximum impact. This also makes all experimentation activities relevant to all levels within the organisation. So, be sure to always keep business objectives in mind.

V2.4 Proces

Mature organisations have generally sorted out their experimentation process, with a detailed process incorporating a step-by-step plan for test (series) from A to Z. For those who processes are not yet mature, there is good news: the basis of a solid process does not always depend on factors such as budget, manpower and tooling. The process steps we have covered in previous chapters are not reliant on this.

V2.5 You don't have to be fully mature for a good experiment

Adult experiments are characterized by using the available resources, at the right time and in the right way. Even with the most expensive tooling, dozens of dashboards, the best developers and large amounts of quality traffic, it is nevertheless possible to set up and run worthless experiments. In other words, you don't have to be fully mature to devise and execute a fantastic experiment (series). Moreover, you can observe and learn from the working methods of mature organisations to set up solid experiments yourself.

Curious about how mature the experimentation programme is within your organisation? Then find out by taking the DDMA CRO Maturity Test.

Closing statement

In this white paper we have highlighted the steps to achieve a well-executed experiment that generates results. However, that wasn't all... Through the two in-depth chapters, we have emphasized the value of structured experimentation series, and demonstrated that experimentation is no longer limited to one department, but must permeate every fibre of an organisation. This enables you to make data-driven decisions within every part of your organisation, and to constantly evolve, whether it's marketing strategies, product development, website design or operational processes.

Through this white paper, we want to inspire you to embrace experimentation as a strategy to transform and continuously grow your business. We challenge you to apply the knowledge and insights from this document to your organisation, and to create a culture of experimentation within which taking informed risks and continuously striving for improvement are the central focus.

Together, we can embrace the benefits of experimentation and break new ground for success. We wish you every good fortune on your journey of growth and innovation.

Have you carried out an exemplary experiment (series) between September 2022 and September 2023, and would you like to earn the ultimate recognition for it? Then submit your case for Experimentation Heroes 2023 and who knows, you might get the chance to show your experiment to the crème de la crème of the field on October 31st.

Colophon

Publisher

DDMA
WG Plein 185
1054 SC Amsterdam
T: 020 4528413
E: info@ddma.nl
W: www.ddma.nl

Do you have any further questions or comments about this white paper? If so, send an e-mail to liekewesselius@ddma.nl

Final editing
Bob Younge, Content Specialist at DDMA

Authors

Bas Jansen
Partner Engagement Manager – gTech Northern Europe at Google

Douwe Möhring
CRO Director at dentsu Media

Frank Overbeek
Business Driver at Sitespect

Joshua Kreuger
Manager Experimentation/ Subscription Sales CRO at DPG Media

Mark de Winter
Director of Product at Clickvalue

Mark van Zuijlen
Global CXO Manager at Jacobs Douwe Egberts

Martijn van Oosten
Conversion Optimisation & SEO at the Nationale Postcode Loterij

Nadesh van Louwerijsen
CRO Consultant at Valantic Nederland

Robin Sellaiyah
Optimization Lead – Health Systems at Philips

ABOUT DDMA

DDMA is the largest trade association for data-driven marketing, sales and service. We are a network of advertisers, non-profit organisations, publishers, agencies and tech providers that use data in an innovative and responsible way to interact with consumers. Through knowledge and advice, we help our members to work in a data-driven and customer-oriented manner, to develop a vision on data use, and to deal with legal changes. We also give our members a voice in The Hague and Brussels, and professionalise the sector by developing self-regulation.

For all DDMA studies, go to: ddma.nl/kennisbank