VERT

VISUAL

EMOTIONAL

RESPONSE

TEST

A test kit to perform
tests on visual design

Last update: March 2025

A part of the research project
Visual Color project

www.Visualcolorproject.dk

Zealand

Developed by
Daniel Uhrskov, © 2025

Version 6.0

INTRODUCTION

QUICK INTRODUCTION

VERT is a rapid and carefully structured test method that evaluates how people react to a visual test medium *('visual stimulus')*. With VERT you can test your visual design choices such as colours, fonts, images and image styles, layout, moodboard and much more.

The aim of VERT is to understand a target groups perception related to the visual material you present for them. This is achieved via a systematic framework consisting of axes with opposite pair of words/adjectives called ***polarity parameters***. With these axes the designer and the test participans establish a common language and a system for the test. The test participans are guided to place their opinion between the two opposite adjectives (*two poles*). It could be: *Is this colour scheme more A) Organic or B) Synthetic, or A) Professional or B) Casual.*

With this systematic approach you don't need to explain a lot for the test participants. In addition, the test often gives you **more feedback** than with a classic 'think-aloud test'. However, VERT and the Think-aloud test are closely related in terms of methodology; *The test provides an immediate response to a visual stimulus.*

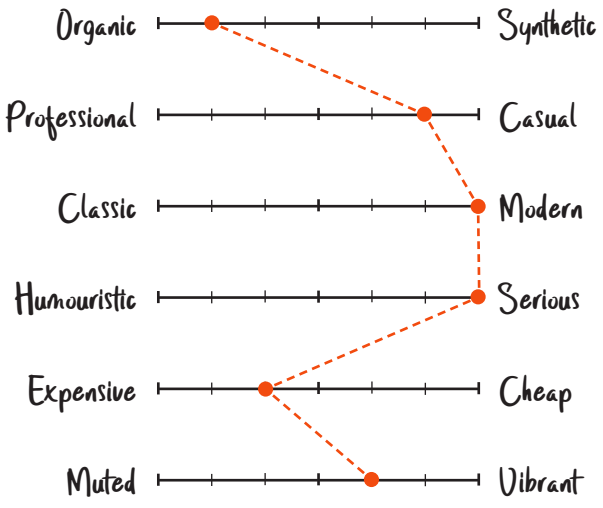


Figure 1: The essence of VERT: the polarity parameters with word pairs and connected dots, transforming it into both a measurable and visually intuitive test method.

WHY YOU NEED VERT?

I have observed - again and again - that it can be difficult for professionals and students to test and gather user data regarding visuals in a design process. In the related research project and literature review (*Visual Color Project*) where VERT has its origin, I describe such observation of a significant lack of effective testing methods tailored specifically to visual design in both educational and professional settings. Students often struggle to obtain actionable feedback on their visual projects, while practitioners in the design industry face similar challenges when trying to validate creative decisions. I always ask my students: *How did you test your visuals*? The answer is often: *I just performed a think aloud test.* The **Think-Aloud method** is a research technique where test participants verbalize their thoughts in real-time while performing a task. This approach provides insight into cognitive processes by capturing immediate reactions and decision-making pathways (1).

My experience is, that existing frameworks for testing visuals, such as the think-aloud test, often produce oversimplified feedback like "*I like it*" or "*I don't like it.*" This lack of nuance limits opportunities for deeper analysis and improvement. *Is the reason that the test participants feel insecure, shy or inadequate when it comes to verbalizing their experience?* It is, for sure, difficult for test participants to articulate their responses to visual elements. *Why?* In the related project (Visual Color Project) and literature review I introduce a theory called *The Missing Link* presenting, that our language is severely limited when it comes to visual verbalization. This can be the reason why we receive simplistic, vague, and unsubstantial responses in our test sessions. I observe that this absence of robust methodologies often leads to stagnation in design workflows. We can call it ***a dead end*** in the design process.

THE FOUNDATION OF VERT

The entire foundation and approach of the research project and VERT stem from the design problem and hypothesis:

It is difficult to put words (language) to one's visual/aesthetic experiences - especially in a test session, both for the designer and the test participant.

Quiet some studies can support this statement. **Think-Aloud testing** can make test participants feel self-conscious and pressured, leading them to alter their natural behavior or provide responses they think are desirable (2). The cognitive effort of verbalizing thoughts while performing tasks can result in fragmented or less authentic feedback (ibid).

VERT will not eliminate these testing factors; The test participans will still feel difficulties and 'stressed'. However, VERT has the advantage of introducing formality and linguistic elements to the test, which can potentially help "facilitate smoother interactions" of the entire test session. It can create a better testing environment and foster a less artificial interaction between the designer and the test participant.

I find a need for a tool that facilitates structured and meaningful feedback on visual designs. I have observed that VERT helped and nudged both designers and test participants to start a conversation about the impact of a visual stimulus.

VISUAL COLOR PROJECT

VERT was developed as part of *Visual Color Project*, a research initiative exploring color's role in design, communication, and user perception. Rooted in an extensive literature review, the project identified a lack of structured methodologies for evaluating visual design choices, leading to the creation of VERT as a concrete knowledge contribution.

While initially centered on color perception and testing, the research revealed a broader **design gap** - a missing framework for systematically assessing visual stimuli. VERT was thus developed as a scientifically grounded yet adaptable tool, refining how designers test and validate visual impact, emotional response, and user perception. VERT functions as both a practical and methodological contribution, offering a structured, repeatable approach

to visual testing. This ensures that design choices are informed not only by intuition but by systematic evaluation and empirical user feedback.

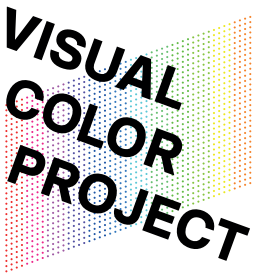


Figure 2: The logo of research project: *Visual Color Project* Www.visualcolorproject.com

BRIDGING THE GAP: THE MISSING LINK IN VISUAL PERCEPTION TESTING

THE CHALLENGE OF ARTICULATING VISUAL EXPERIENCES

Designers frequently encounter challenges when attempting to translate visual aesthetics into verbal language, a phenomenon Marty Neumeier terms "The Brand Gap"—a disconnect between visual rhetoric and linguistic articulation (3). Similarly, Anne Mette Hartelius describes a "design gorge," highlighting the theoretical gap in how we understand and communicate visual meaning (4). This absence of a common vocabulary for visual perception complicates both the design process and user testing, where participants struggle to articulate their intuitive responses to visual stimuli. The VERT method directly addresses this issue by providing a structured, semi-linguistic framework that facilitates the articulation of aesthetic experiences. By introducing structured adjective axes, VERT creates a guided yet flexible test environment where users can express perceptions through a comparative and systematic approach rather than relying solely on free verbal expression. While this does not entirely eliminate the complexity of verbal overshadowing - the cognitive interference that occurs when individuals attempt to describe non-verbal stimuli in words (5) - it provides a structured aid to bridge this cognitive gap.

BEYOND LANGUAGE: QUALIA AND THE NON-VERBAL EXPERIENCE

Cognitive science suggests that not all knowledge can be fully captured through language. The concept of qualia, as introduced by Frank Jackson, describes subjective experiences that cannot be entirely communicated through words (6). Color perception is a prime example of this limitation, as seen in studies on linguistic relativity in color naming, where certain cultures distinguish fewer color categories, yet their perceptual experience remains unaffected (7). Similarly, Aaron Fine argues that color should not be understood solely through linguistic structures but also through emotional and sensory intuition (8). This raises an important question in visual testing: How do we measure and communicate something that is inherently beyond language? VERT does not attempt to reduce visual perception to mere data points; instead, it reserves space for subjective experience by providing structured yet open-ended ways for participants to indicate their reactions. The method's reliance on bipolar adjective scales nudges participants towards articulating their experience without imposing rigid linguistic constraints. This creates a semi-structured dialogue between test facilitator and participant, where language aids but does not dictate the interpretation of visual elements.

EMOTION VS. FEELING: THE DUAL PROCESSING OF VISUAL PERCEPTION

Joseph LeDoux (1998) highlights that emotional responses can be processed both unconsciously through the amygdala and consciously through the cortex, suggesting that human reactions to visual stimuli operate on multiple cognitive levels (9). This distinction between emotion (pre-verbal, instinctive) and feeling (consciously articulated, linguistically shaped) is particularly relevant in the VERT framework. Martin Skov (2007) and Morten Kringelbach (2007) argue that aesthetic experiences engage both these cognitive pathways, meaning that a successful evaluation method must account for both rapid, instinctive reactions and their later verbalized interpretations (10+11). Traditional think-aloud protocols primarily capture conscious, articulated reflections throug language, potentially missing the more immediate, visceral responses that influence perception. VERT, by contrast, experiements and tries to capture the more immediate responses through the polarity axes - that doesn't demand the same linguistic processing - while also leaving room for later reflection and verbal articulation through qualitative follow-up questions or moments.

NUDGING, NOT DICTATING: THE ROLE OF ADJECTIVE AXES IN VERT

A fundamental principle in designing effective user tests is avoiding leading questions or priming effects, which can bias participant responses. The VERT method is built on the principle of ***nudging*** rather than directing test participants. By presenting pre-defined adjective axes, the method guides users toward structured articulation without prescribing a singular, fixed interpretation of the visual stimulus. This is a crucial distinction: rather than dictating a response, **VERT serves as a catalyst for cognitive reflection**, allowing participants to frame their reactions within a structured yet flexible evaluative system. The inclusion of blank axes for participant-generated adjectives further reinforces this openness, ensuring that the test remains adaptive rather than restrictive.

CONCLUSION: A COMPREHENSIVE FRAMEWORK FOR EVALUATING BOTH CONSCIOUS AND NON-VERBAL EXPERIENCE

The challenge of evaluating visual stimuli lies in its inherently non-linguistic, experiential nature. Traditional user testing methods often fail to capture the immediacy and depth of visual perception, either because they rely too heavily on verbal articulation (as seen in think-aloud protocols) or because they reduce complex experiences to numerical ratings - This is the empirical observation and thus the hypothesis and entry point assumed in this research project. VERT offers an alternative by creating a structured yet open-ended test environment that balances quantitative measurement with qualitative insight. By facilitating the articulation of non-verbal impressions through structured axes and allowing for participant-driven input, the method effectively bridges the gap between tacit visual knowledge and explicit verbal communication.

THE DEVELOPMENT OF VERT

My inspiration for developing VERT comes from the so called BERT. Below is what the BERT test originally stands for:

> **BERT: BEHAVIORAL, EMOTIONAL, RESPONSE TEST**

It's difficult to date the BERT method, but it is primarily used in marketing. However, my research revealed limited academic resources validating or detailing BERT's origins, with most information stemming from practical sources such as blog-style introductions on *UX for the Masses* by Neil Turner (12), *UX Design* by Sherwin Pollack (13), and *Clearleft* by Harry Brignull (14). While these resources provide valuable insights into BERT's application, they lack a rigorous theoretical foundation.

Inspired by George Kelly's *Repertory Grid Technique* (1955) (15) and J.A. Russell's Circumplex *Model of Affect* (16), I developed the VERT kit. Kelly's framework, which captures subjective evaluations using bipolar scales, aligns

closely with BERT's approach to measuring user sentiment. Similarly, Russell's circumplex model maps emotional states across dimensions of arousal and valence, providing a nuanced way to analyze affective responses. These theoretical underpinnings informed the structure and metrics of VERT, enabling it to evaluate both emotional and rational dimensions of visual design. Currently, there are a few online sources that provide inspiration on how to work with the axis, but no research, experiments or formal evaluation of the method.

VERT is designed to 'measure' and gather data from test participants' immediate, impulsive reactions to **visual stimuli**, whether that be an interface, layout, color palette, or other visual elements. In developing this new test method, I want to specialize BERT for testing visuals, which is why I named it VERT:

> **VERT: VISUAL, EMOTIONAL, RESPONSE TEST**

SCIENTIFIC METHODS IN DEVELOPING VERT

The VERT kit has undergone multiple iterations, informed by observations of approximately **200 design students** using it and providing feedback through questionnaires and oral feedback sessions. These refinements have focused on ensuring that VERT remains intuitive, accessible, and easy to implement. The structured yet adaptable nature of VERT allows its application across diverse design contexts, from assessing branding elements to optimizing user interfaces. By grounding its methodology in both theoretical frameworks and practical usability, VERT offers a systematic approach to visual testing, providing actionable insights that support iterative design processes and enhance design decision-making. The development of VERT was rooted in **three primary scientific methods** (*methodological triangulation*), all aligned with an *explorative design approach*. The empirical data documented in the separate journal: *Empirical Foundations and Explorative Development of VERT*, provides a detailed explanation of the development of the VERT kit (17). In the following section, I will present the three key highlights from this process.

First, a pilot study was conducted, involving two design students who tested an initial version of the kit in a project about visual identity. The design process leading to this pilot study was grounded in my literature review and developed frameworks. Initially, I dedicated time conducting further research and interviews as well. I developed a prototype of the VERT Kit, ensuring that it had sufficient value and practical application to be tested in-depth. When that prototype was ready, the two students implemented it in a project and tested it. Pilot studies, as emphasized by Van Teijlingen and Hundley (18), are critical for identifying potential challenges, refining research tools, and enhancing feasibility. Therefore I decided to facilitate this pilot study as a qualitative test session before further development of VERT. This approach provided direct input from relevant stakeholders and allowed me to witness how the tool was applied in practice. While the prototype showed potential, there were numerous areas for improvement and critical feedback points. The key objectives of the pilot study was to gather early-stage feedback that could shape and refine VERT. Observations and feedback from the students made it possbile to do adjustments, particularly in terms of usability and clarity. Even though I got a lot of critical feedback, I also, immediately observed the potential and the value of VERT.

Second, a session utilizing *cultural probes* offered valuable qualitative insights. Further developed versions of VERT (version 2.0) were distributed to 18 groups of design students, who were asked to use the tools freely and document their usage. They got a short manual, the test pages and some markers. As Gaver, Dunne, and Pacenti (19) outline, cultural probes facilitate open-ended exploration and user-driven feedback, making them ideal for capturing diverse perspectives on tool usability. This approach revealed key areas for improvement, such as enhancing the kit's flexibility, neutrality, and graphical design, while also generating critical oral and written feedback. I got the 'executed' VERT kits back from the students and could thereby analyze how VERT was used. I facilitated a feedback session with all the students and I also collected data in a survey.

Third, a structured comparative testing session was conducted with two classes of design students: 1 test with Class A (October 2023 - *VERT 3.0*) and 1 test with Class B (October 2024 - *VERT 4.0)*. In the year between the tests I further developed VERT with focus on the result from Class A. **The comparative aspect involved comparing VERT with a more traditional test method** to understand its effectiveness, advantages, and limitations. This comparative test was a mix of observations, oral feedback and collecting detailed answers and rating from the students. The goal of the test was to extract qualitative insights and data from a session in which I closely observed the use of VERT. The objective was to gain valuable insights, make final but crucial adjustments, and conduct a conclusive evaluation of VERT's value within the design process.

The students were asked to test a mood board, including a color scheme, using both a conventional think-aloud test method and after VERT. It should happen with two different test participants. The think-aloud protocol, as described by Ericsson and Simon (20), is widely used to capture participants' verbalized thought processes during a task. However, it was this project thesis and expectation, that its application in visual testing often results in overly simplistic feedback. And when I evaluated the 2 test methods with the students it was clear, that this was the case. VERT enabled more detailed, structured feedback, promoting richer conversations between designers and test participants. By incorporating the bipolar axes the participants was nudged to share more feedback. This approach helped the students to capture specific insights about emotional resonance and the effectiveness of their visual choices. But it also helped the test subjects to gain a rapid, impulsive language about the visual stimuli and it facilitated the development of language, streamlined the dialogue, and enhanced the quality of feedback. I observed that the test participants 'opened up' and provided more feedback and that there was plenty of oral feedback not only related to the axis and adjectives as well.

Through these 3 stages of developing VERT, the iterative and exploratory nature of the VERT Kit became evident, reflecting the principles of explorative design as described by Sørensen, Mattsson, and Sundbo (21). By integrating structured comparative testing with participant feedback, the VERT Kit evolved into a versatile and robust tool for visual design evaluation, addressing both the need for detailed analysis and the flexibility required for nuanced feedback. The design process of VERT contains the principles of explorative design, characterized by cycles of design, observation, reflection, and refinement (ibid). This process parallels the IDEO Design Thinking (22) framework, which emphasizes user-centric, iterative innovation to solve complex problems. By continually iterating, the research integrated empirical insights, interdisciplinary methods, and practical adjustments to create a more robust tool for visual design evaluation. Though the feedback and development also led to some critisism of VERT that I will now describe:

CRITICISMS AND FURTHER DEVELOPMENT

Criticism of VERT emerged during the development and the students offered valuable insights to refine VERT. For instance, some expressed concerns that the structured axes in VERT limited test participants' ability to articulate their full emotional and linguistic responses freely. *What if a non-included emotion was relevant? Didn't the axis guide the test participant in a certain direction?*

The comparative testing of VERT versus the traditional think aloud test demonstrated the practical benefits of the VERT Kit. However, the session also revealed areas for improvement. For example, some students observed that the kit's structured approach could inadvertently restrict the spontaneity of participants' responses. These critiques informed subsequent iterations of the VERT Kit, including the addition of more flexible components and a recommendation to incorporate unstructured feedback sessions alongside the formal evaluation framework. These adjustments were instrumental in creating a tool that balances structure with openness, ensuring nuanced and objective test results while maintaining the kit's utility for systematic analysis. Though it was clear, that

there were plenty of moments with 'free and open' conversation in the test sessions with VERT. But the importance and facilitating of these conversations should be mentioned in the manual.

To address these concerns, guidelines were added to this manual of the VERT kit on *how to facilitate the test, allowing space for open-ended feedback where test participants could freely express their thoughts and feelings without constraints.* Furthermore, 2 empty axis was added on the test sheet for the test participants to fill out with their own adjectives. Additionally, the feedback highlighted the importance of ensuring the objectivity of test sessions, avoiding any influence from designers that could bias participants' responses.

IS VERT QUALITATIVE OR QUANTITATIVE?

VERT incorporates **both** qualitative and quantitative aspects, making it a flexible tool for visual testing.

The **qualitative dimension** is reflected in the test's ability to foster deeper understanding of a test participant's emotional and linguistic responses to visual perception. Designers aim to empathize with their audience, and in this context, VERT can be supplemented with think-aloud techniques, allowing designers to ask follow-up questions and explore specific reactions in more depth after the test participants have completed the axis-based evaluation. At the same time, VERT is **primarily a quantitative method**, as its core function is to measure, calculate averages, and compare test results against an intended design objective. The more participants included in the test, the more solid the data foundation becomes for drawing conclusions. However, unlike large-scale surveys, VERT is not necessarily designed to function as a mass-data collection tool - although it can be adapted for that purpose as well.

Ultimately, VERT is highly **customizable**, allowing designers to tailor the method to their specific needs within the design process, balancing qualitative insights with quantifiable validation. In this manual and the example included, the test is balanced, incorporating elements of both qualitative and quantitative insights.

DIGITALIZATION

Although VERT is currently a analog tool, its potential for digital transformation is considerable. A digital version could integrate advanced analytics, real-time feedback, and dynamic interfaces to enhance its usability. Based on my design process with VERT and the numerous observations and tests I have conducted, it is evident that the test execution itself is most effective when performed in an analog, manual and tangible format. This approach clearly separates the test sheet (axes) from the visual test media, facilitating a more intuitive understanding and distinction between the test elements. Therefore, my recommendation is to maintain the test execution on paper or a 'digital paper' format to preserve its clarity and usability.

While the test evaluation can be conducted digitally using Excel (**a template is available for download**), there are also opportunities to further refine and visualize test results digitally. In its present form, VERT remains more analogue and tangible. Even the evaluation can be done by hand.

THE INTENTION WITH VERT

VERT is not merely a data collection tool - it is a facilitator of dialogue and a **catalyst** for deeper articulation in visual testing. While the structured axes provide a framework for measuring responses, their value lies in nudging participants toward verbalizing their intuitive reactions. With the axes VERT reduces the cognitive burden of articulating visual impressions from scratch, making it easier for participants to express their perceptions spontaneously. The unstructured feedback; verbal reflections, spontaneous comments, and test participant explanations, remains equally significant. Recognizing this, VERT incorporates guidance on how to interpret and document qualitative responses that emerge during testing. This ensures that both structured data and open-ended insights contribute to a more nuanced understanding of how visual elements are perceived.

PARAMETERS

These are the flexible parameters that can be adjusted and optimized to suit your unique test:

- Amount of polarity parameters
- Amount of test participants and test sessions
- Amount of test medias and the test setup
- The amount of tests during the design process - maybe you will do several VERT test separated from each other

THIS MANUAL

This manual serves as a step-by-step guide to six stages of the VERT methodology, providing structured inspiration for integrating the test into your design process. This manual includes a practical example of VERT, test sheets templates and adjective pools. Additionally, an **Excel template is available for download**, to facilitate efficient analysis and organization of test results.

VERT GLOSSARY & PARAMETERS

AXIS

The horizontal bands that form the primary structural framework of VERT.

MEASUREMENT POINTS

Each axis contains seven evenly spaced marks, ranging from -3 to +3. 0 representing the neutral midpoint.

PAIRS OF WORDS

Contrasting word pairs with opposite values.

POLARITY PARAMETERS

The axes combined with their respective pairs of contrasting words, establishing the measureable dimension.

DOTS

Markers that represent recorded responses on the polarity parameters.

CONNECTING LINES

Visual lines that link the response dots across polarity parameters. These makes the test result more visual and easy to interpret. The use of color is beneficial for differentiation.

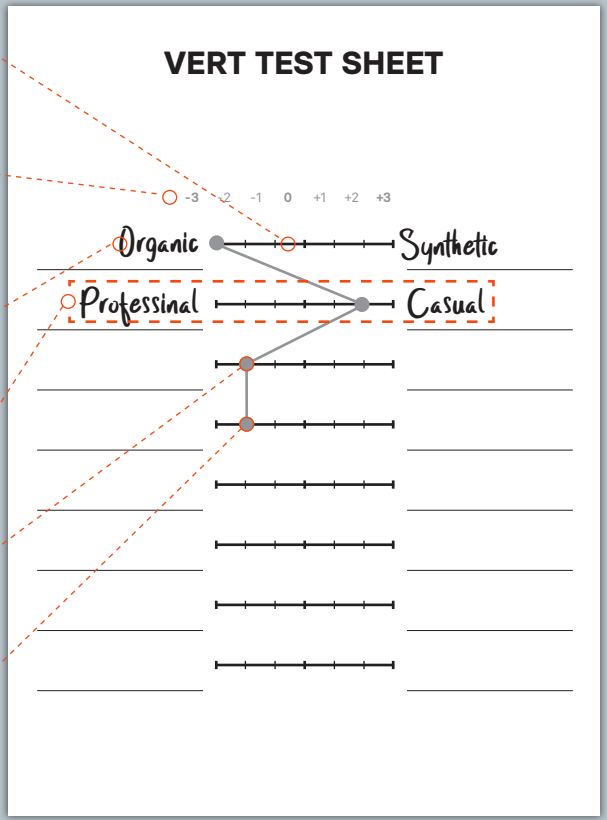


Figure 3 (right): Infogrpahic showing the systematic and graphics of VERT.

6

The 6 steps
+ Quick manual

PREPARATION

1) Visual Test Medium

2) Test Participants

3) From Words to Polarity Parameters

4) Test Sheets + Visual Hypothesis

EXECUTION

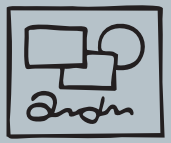
5) Test Setup and Conduction

MONITORING

6) Visual Evaluation and Further Process

QUICK MANUAL

1



DESIGN VISUAL TEST MEDIA

Plan and prepare the visual material you want to test. Make sure it accurately represents the final design.

2

FIND TEST PARTICIPANTS

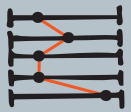
Decide who to test. Think about your target group and select a reasonable number of participants.

3

CREATE AXIS

Choose meaningful word pairs that represent opposing values. Make sure they align with the intended perception of your test media.

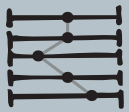
4



DEFINE THE INTEDED VALUES

Place dots on the axis based on the intended value of design. Connect the points with a colored, vertical line. Save this hypothesis sheet!

5



COLLECT ANSWERS

Present your test medium to the participants. Allow them to mark dotss on the empty axes. Remain open to oral feedback and conversational elements, as these may provide valuable insights. Conduct more tests to ensure more comprehensive findings.

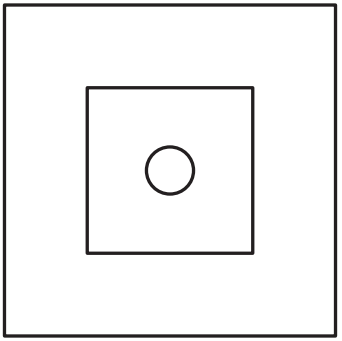
6

COMPARE

Compare the participants' responses to your hypothesis sheet. Do they match? Are there any differences? If so, what adjustments might improve alignment?

Step 1

What should be tested?



Visual Test Medium

WHAT SHOULD BE TESTED AND WHY?

This first step involves creating the actual visual test media; *The object that you want to test.* VERT was developed with a focus on colors and color schemes as its central theme, but in principle, **all visual products can be tested**. However, the test is specifically designed for more or less "static" visual sensory impressions, as interactions are likely to divert focus away from these impressions. For instance, while it is possible to test the emotional impact of a website using VERT, it is not recommended to take the user on an extended journey through the website or incorporate more complex usability testing questions, as these may detract from the focus on the visual aspects.

Examples of what can be tested includes:

- Colors, color combinations, and color usage
- Typography and typographic hierarchies
- Moodboards
- Layout and design principles, visual weight, shapes, and graphic elements
- Motion graphics
- Printed materials

REPRODUCE AS ACCURATE AS POSSIBLE

The visual test media is the material you show to your test participants. When preparing the visual test media, it is important to consider how you can thoughtfully represent the visual design choices in alignment with the final product. **If you do not have the actual product or concept yet, try to reproduce the visual elements as accurately as possible**. *How can you best calibrate the elements?* For example: print your test media if the final product is a printed matter. Similarly, if you are testing the appearance and layout of a website, it is best to present the site on the type of screen the target audience will typically use to interact with it. Consider to use a graphic mock-up to present your design. If it's packaging, you could consider to implement your style on a product mock-up. If you know the background color of the context you are working with, it is crucial to use that specific background color behind a color scale, font, or other elements being tested. **Consider that all parameters play a role**. Imagine that you test a font. It would not only be the letter design that is tested, but also the styling, the font hierarchy, the size, the font color, the spacing etc.

If you are testing colors: Do all the colours take up the same amount of space? Or is there a larger area of one colour in the design you've created? Some colours should take up more space than others on the test media. Some colours may play a small role - maybe in details. But they can still have a big impact. One example is call-to-action (CTA) colours, which are a kind of eye-catcher or strong "spot" colours. If your design uses a lot of beige and a little bit of black and red, then the test media should reflect this too. You cannot simply show a color scheme of 3 identical squares in beige, black and red.

To conclude, **it is not only about the visual elements you choose, but also the elements' visual 'weight' in a 'visual hierarchy' that should be reflected on your test media**. Your test media could be a product, a moodboard, a logo etc. It should be a clear and framed grpahic 'artifact' and it should not include a lot of shift, interactions and dynamic changes.

IF YOU DON'T KNOW THE PRODUCT OR MEDIA

It can be tricky to decide and argue for your test medias appearance. Especially if you don't know how and where the visuals or colours are applied. But you should consider the test media as a kind of *board* where you can experiment and test different visual compositions that could later be applied to your design. This *board* is explained further in the next block.

CREATE A MOODBOARD

If you aren't sure how to test an idea or what you need to show, I recommend to create a 'board', that represent the visual style you want to test and potentially implement. This might be a so called *moodboard*. The moodboard can contain many visuals such as colors, fonts, graphics, photos, shapes etc. Moodboards are an effective tool for both individual designers and teams to visualize a design direction by seeking inspiration, collecting materials, and curating a potential visual style. This style is materialized on a board, which functions much like a pinboard. On this board, the designer selects and excludes elements that, together, form a cohesive direction called a moodboard. I observe that Moodboards are widely used in the industry due to their simplicity and low cost of execution. However, there is limited academic literature on the subject.

Anne Mette Hartelius describes moodboards as a technique for generating ideas, where design intuition is employed to develop visual tonalities (23). Through moodboards, designers explore hypothetical directions for design concepts. The aim is to narrow down the potential visual tonality of a brand - aesthetically speaking. A moodboard should consider brand personality, brand narrative, tonality, target audience personas, cultural insights from both sender and receiver perspectives, and brand drivers (ibid).

Within the context of VERT, it is crucial that the moodboard is self-explanatory (ibid). The conent of your moodboard (images, fonts, colors ect.) must independently convey the mood and feeling without being aided by descriptive text or guiding formulations. Only a title and a brief description should accompany the board.

Morten Rold briefly addresses the tool in his book *Interfacedesign*, describing a moodboard as a *collage consisting of images, words, colors, and shapes designed to visualize an expression and the feeling intended for a design* (24). However, one of the challenges with moodboards is their abstract nature, which can make it difficult for clients to envision how the moodboard translates to a specific interface (ibid). Similarly, in the context of VERT, test participants may find it challenging to imagine how and where a moodboard will be applied concretely.

That said, it is important to note that VERT can also be used to test specific products or websites, as long as they are relatively static and the test remains focused on the visual aspects. Moodboard can be used, if the product is not developed yet or if you want to work on the visual style separately. For more information and visual inspiration on moodboards, I recommend these books: *Nordic Moods - A Guide to Successful Interior Decoration* by Katrine Martensen-Larsen (25) and *Maison: Parisian Chic at Home* by Ines de La Fressange and Marin Montagut (26). Moodboards, as presented in these books, are particularly prevalent in the domain of interior design, likely originating from this field's methodological approach to creating atmospheres in spaces and interiors. This could explain the lack of academic literature on the subject. Nonetheless, moodboards are equally applicable creating interfaces, marketing materials, packaging, and more.

HOW MUCH TO TEST AND HOW MANY TEST MEDIA?

Though you can test several and different test medias - also in the same session. It is, actually, good to have some different visual test medias to test various design directions. Not too many though as the test participants can experience what I call 'visual fatigue' or 'perceptual overload' like other studies do. This can be a barrier in user testing within UX (27). It's better to perform fewer and more quality-based tests on each person and 'evolutionarily develop' the same visual test media over time. VERT is particularly effective for conducting A/B split tests, allowing designers to compare different visual design alternatives systematically. This approach aligns with established A/B testing methodologies, where two versions of a design are compared to determine which performs better in terms of user engagement and satisfaction (28).

If you test more media, be sure to 'clear' and separate each session from each other. However, it is essential to avoid overloading participants with too many variations in a single session, as

this can lead to 'visual fatigue' or 'perceptual overload,' potentially hindering the effectiveness of the test (26). Therefore, limiting the number of variations and conducting multiple sessions can yield more reliable insights.

Finally you can execute VERT tests in both 1) the initial stages of the design process, long before the final product, to gather early and valuable feedback. And 2) later in the design process, close to the final product, to further validate that a visual concept functions as intended.

SUM UP AND PRACTICAL TAKEAWAYS

- Design your test media with purpose.
- Ensure the visual test media accurately reflects the final product's intended look and feel. If the actual product is unavailable, create a representation that best simulates its application. Consider all parameters - background color, font size, spacing, and hierarchy - so that your test media presents a realistic and meaningful composition. If the design will be printed, present it as a printed test media. If it's digital, test it on the type of screen your target audience will use.
- Avoid perceptual bias: Keep the test focused on the visual experience. Avoid adding elements that could bias the test participants, such as words, symbols, or instructions that guide their perception.
- Use moodboards for conceptual testing: When testing early-stage designs, moodboards can be an effective way to gauge emotional responses to color, typography, and general visual direction.
- Limit the number of variations: Presenting too many test media versions can lead to perceptual overload. Fewer, more structured tests provide deeper insights while allowing for iterative refinements.

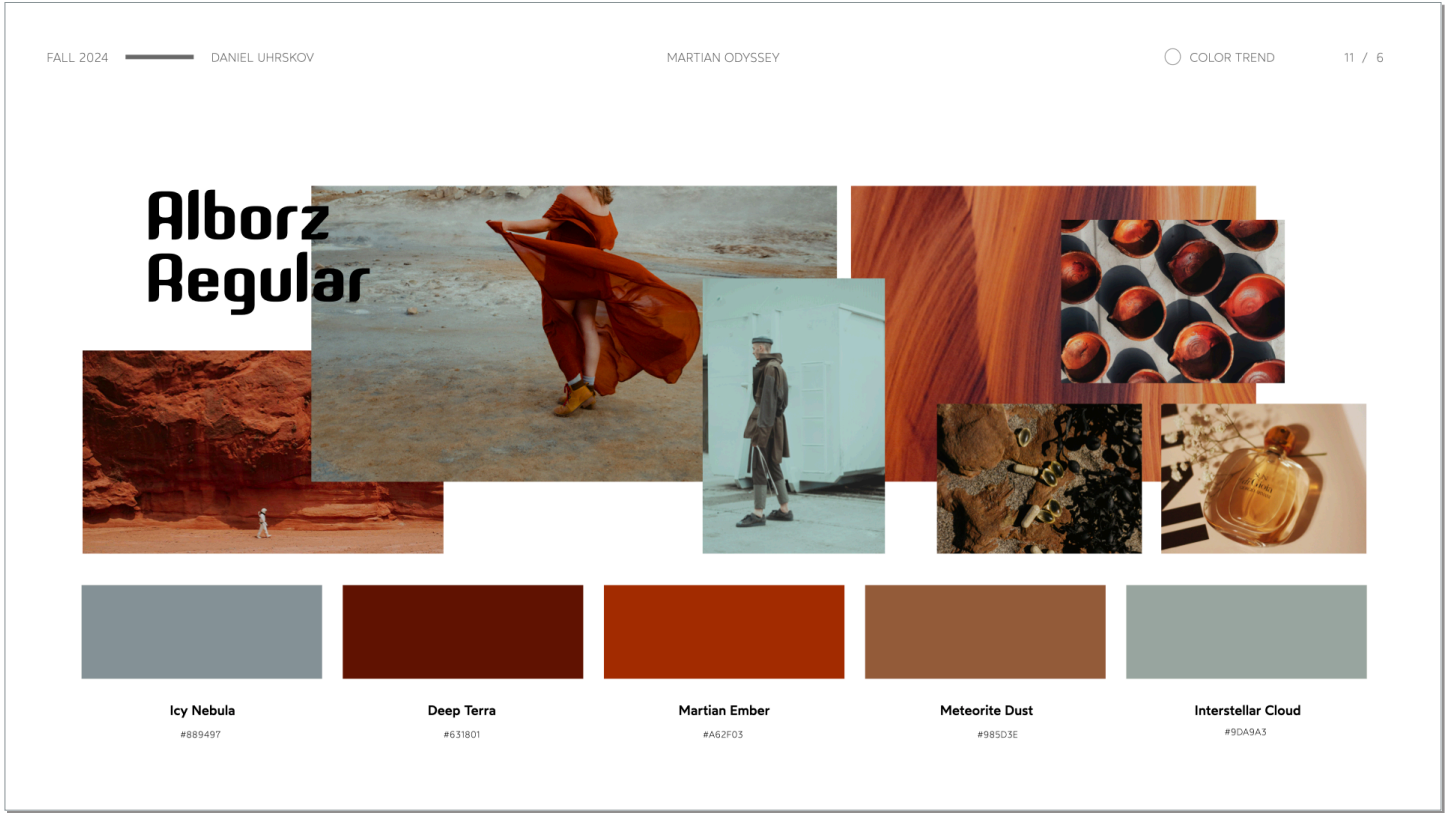
Practical Example

EXAMPLE THROUGHOUT THE MANUAL

The moodboard in the next page serves as the **practical example** of VERT in this manual. At the end of each step, you will find this example illustrating how the process could be conducted with this moodboard as visual test medium. This allows you to follow a practical application of the test method, providing a clear and structured understanding of its implementation.

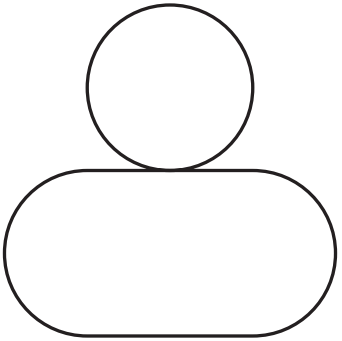
THIS IS A MOODBOARD

Figure 4: Martian Odyssey, 2024, Developed by Daniel Uhrskov Hilleberg



Step 2

Who should be tested?



Test Participants

FINDING RELEVANT TEST PARTICIPANTS

Finding the right test participants is crucial for obtaining reliable and valid results in visual design evaluations. Participants should closely represent the target audience of the product/ concept/brand etc., considering factors such as age, gender, cultural background, and economy or more important the specfiic *life stage* (find the following block: *Life Stage Perspective*). Utilizing insights from target group analyses, data collection, or desk research can aid in identifying suitable participants.

This research project and VERT is relying on the standards within user testing and target group analysis in UX-design, such as Design Thinking or other iterative design frameworks. Therefore I will not elaborate further on this topic, as it is out of the scope of this project. Some terms that is relvent will be mentioned in the following section.

PERSONA & SCENARIO

- Persona: A fictional character representing a specific user type based on real data and research. It includes demographic details, behaviors, needs, and pain points (28).
- Scenario: A narrative describing how a persona interacts with a product or service in a specific context to achieve a goal. Scenarios help visualize user experience and pain points.

In the context of VERT, recruiting test participants who closely match the defined personas is essential. Personas serve as a guiding framework for selecting representative users, ensuring that the feedback collected is relevant and aligned with the target audience's perceptions and expectations (ibid).

USER JOURNEY

- A visual or narrative representation of a user's interaction with a product or service over time.
- Includes touchpoints, emotions, motivations, and barriers experienced at different stages.
- Helps identify pain points and opportunities for improvement in the user experience.

User journey frameworks are essential for understanding the experiences that personas or target groups have with a product, service, or brand. They help identify touchpoints and moments of tension or friction, enabling designers to recognize areas for improvement (29). By mapping these interactions, designers can strategically integrate tools like VERT into specific stages of the user experience and design process, ensuring that visual evaluations align with key decision points.

VALUE PROPOSITION CANVAS (VPC) & EMOTIONAL TRIGGERS

- Value Proposition Canvas (VPC): A tool developed by Osterwalder et al. (2014) to align a product's value with user needs (31).
- Emotional Triggers: Psychological factors influencing user decisions and engagement.

The VPC helps align a product's value with user needs, ensuring that design decisions resonate with the intended audience (ibid). Similarly, emotional triggers - the psychological factors influencing user decisions - play a crucial role in visual perception and branding. Understanding these elements allows designers to identify the right test participants for VERT, ensuring that feedback is gathered from individuals whose perceptions and emotional responses align with the target audience.

SEGMENTATION MODELS

Methods used to categorize target audiences based on shared characteristics to tailor marketing and UX strategies.

Key Segmentation Models:

- Demographic Segmentation (age, gender, income, education).

- Psychographic Segmentation (lifestyle, personality, values).
- Behavioral Segmentation (buying habits, brand loyalty).
- Geographic Segmentation (location-based targeting).

Segmentation models are essential for categorizing target audiences based on shared characteristics, ensuring that user testing reflects the right demographic, psychographic, behavioral, or geographic groups. By applying segmentation frameworks, designers using VERT can refine their test participant recruitment to align with the most relevant audience segments.

LIFE STAGES PERSPECTIVE

Rather than focusing on age, gender, or location, considering life stages provides a more relevant approach to understanding user perception. The key question is: *Where is the test participant in life, and does this stage align with the target group?* A 27-year-old woman could be a student, pregnant, or a new mother, each stage influencing her consumption and perception differently. As Kongsholm (2020) notes, life stage matters more than age in shaping user behavior (32).

Finding representants from a chosen, relevant life stagewould be valueable in the test of VERT. Including participants from a relevant life stage ensures that the insights gathered through VERT reflect not just demographic data but also the contextual and psychological factors that influence perception and decision-making.

RECRUITMENT STRATEGIES

When the target audience is hard to reach, alternative recruitment methods are essential. Leveraging personal networks, online platforms, and user communities can help identify suitable participants. Even if participants do not match the exact target group, they should share key characteristics relevant to the study, ensuring validity while managing recruitment challenges.

Use some of the mentioned methods in this sections to choose and get in touch with the target audience.

DIGITAL VS. MANUAL TEST EXECUTION

Although VERT is designed for manual execution in a test setup, digital distribution can be useful when participants are difficult to access or geographical constraints exist. However, in-person testing allows for better observation of responses and richer qualitative insights. The choice depends on the study's objectives and available resources.

HOW MANY TEST PARTICIPANTS?

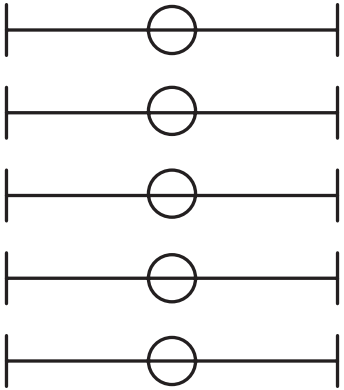
For details on sample size recommendations, refer to Step 5: '*What is the Optimal Number of Tests?*' on page 38, which also addresses the number of required test participants.

SUM UP AND PRACTICAL TAKEAWAYS

- In summary, careful selection and recruitment of test participants, considering both demographic factors and life stages, are vital for obtaining meaningful insights in visual design evaluations.
- Make sure that your target groups and/or personas are represented among the test participants considering demographics, cultural background etc.
- Life stages can provide a deeper perspective beyond traditional segmentation, offering more insight into user behavior in the context of visual perception.
- Digital testing is possible but should be used selectively, as manual execution ensures better observational data.
- A well-selected participant group enhances the validity and relevance of VERT results in visual design evaluations.

Step 3

What should be tested?



From Words to Polarity Parametres

GENERATING ADJECTIVES AND CREATING POLARITY PARAMETERS

A structured process is essential for selecting the right adjectives and polarity parameters to ensure meaningful test results. This step consists of three phases: brainstorming words, forming polarity pairs, and selecting the most relevant parameters for testing.

STEP 1: BRAINSTORMING RELEVANT WORDS

The first step is to *generate a broad list of adjectives* that describe potential emotional and behavioral reactions to the visual test medium. These adjectives should be simple, emotionally resonant, and easy to understand, ensuring that test participants can intuitively relate to them.

Brainstorming is a key ideation technique widely used in design thinking to generate diverse ideas before refining them into structured concepts. It encourages free-flowing, non-judgmental idea generation (33). In this case brainstorming should allow a broad exploration of potential adjectives that describe emotional and behavioral responses to the visual test medium. This is how you can brainstorm keywords:

- **Set a clear goal** – The objective of this brainstorming session is to identify adjectives that **capture the intended visual perception**, emotional reaction, and brand associations. It could be: Exclusive, Professional, Quiet.
- **Generate a wide range of words** – Write down as many descriptive words as possible without filtering or evaluating them yet. Consider aspects like color harmony, usability, aesthetics, and mood.
- **Expand beyond adjectives** – While adjectives are the most relevant, nouns or verbs can also provide meaningful insights (e.g., Luxury – Budget, Attracts – Repels).
- **Refine the list** – After the brainstorming phase, review the adjectives, remove redundant ones, and prioritize those most relevant for testing. If possible, conduct the session with colleagues or stakeholders to gain multiple perspectives.

STEP 2: FORMING PAIRS OF OPPOSING WORDS (POLARITY PARAMETERS)

Once the most relevant words have been identified, the next step is to organize them into polarity pairs - two opposing values that create a measurable scale for participants. The goal is to determine how test subjects perceive the design on each parameter.

- Identify opposites: For each selected word, find a meaningful opposing term (Exclusive > Discount, Professionel > Humouristic, Quiet > Vibrant).
- Ensure relevance: The polarity pairs should align with what you are testing - whether it's aesthetic qualities, brand perception, or user experience.
- Avoid redundancy: Each pair should measure a distinct perception, avoiding overlap with other pairs.

STEP 3: SELECTING FINAL POLARITY PARAMETERS

Not all polarity pairs need to be included in the final test. It's important to prioritize and refine the selection to maintain clarity and prevent cognitive overload for test participants.

- **Select 6-8 key polarity parameters** that are most relevant to the test's purpose.
- Ensure there is space for additional input: Test participants can define up to two additional custom polarity parameters if they feel something is missing. There is already reserved space for them on the test sheets.
- Consider what you want to confirm: What is the most critical aspect of the design that you need feedback on?

ADJECTIVE POOLS FOR INSPIRATION

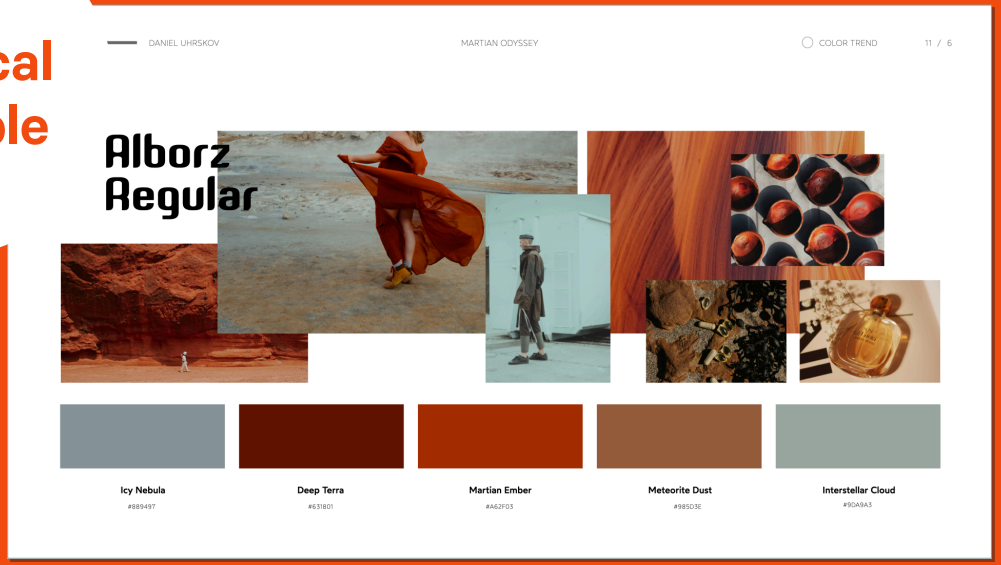
To assist in structuring polarity parameters, predefined adjective pools have been developed for different visual domains. They can be found at the end of this manual. These serve as a starting point but can be customized to fit specific testing needs. The included pools are:

- **Colors** (e.g., Vibrant – Muted, Harmonious – Clashing)
- **Typography** (e.g., Classic – Modern, Legible – Confusing)
- **Web Design** (e.g., Minimalistic – Overwhelming, Clear – Cluttered)

SUM UP AND PRACTICAL TAKEAWAYS

- Select Clear Adjectives: Choose simple, emotionally resonant, and contextually relevant adjectives. Review them before finalizing.
- Create Polarity Parameters: Combine adjectives into opposing pairs that align with the test's purpose. Keep it under 8 pairs to ensure focus and reliability.
- Use Adjective Pools: Predefined adjective pools for colors, typography, and web design are available as inspiration. Consider customizing them to match specific testing goals. Find them at the end of this manual.

Practical Example



PICKING THE RIGHT ADJECTIVES

The words on the right have been carefully selected to test the example moodboard (Martian Odyssey) in a real case scenario. These words were chosen based on the **intended vision** for the moodboard, which aimed to create a **organic, down to earth, modern/futuristic, serious, exclusive and vibrant feel**. By aligning the test axis with the design vision, the evaluation becomes more structured and meaningful.

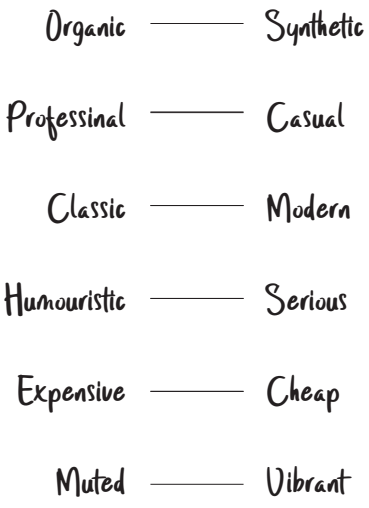
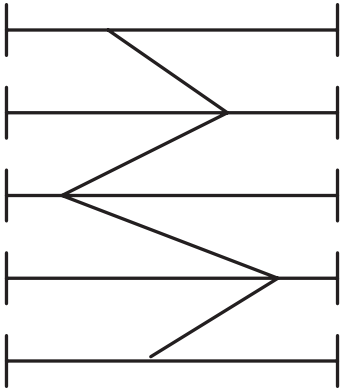


Figure 5: Selected word pairs representing the desired design values, with their opposing counterparts as a reference.

Step 4

What do you aim?
What is your vision?



Test Sheets + Visual Hypothesis

TRANSFERING ADJECTIVE PAIRS

Transfer the selected word pairs (from Step 3) onto the **hypothesis sheet**, where they will serve as **polarity parameters on the axes**. You can find this sheet at the end of this manual. Each axis consists of seven measurement points, with **0 as neutral**, while **-3** and **+3** represent the extremes of each polarity parameter.

Leave at least **2 axes empty** for test participants to fill out if they experience emotions or responses that are not represented. This 'reserved' space ensures greater objectivity and openness in the test method.

MAKING IT MEASURABLE

This step makes your selected words measurable and establishes the fundamental structure of the test. Of course, you could simply extract data from the test using measurements ranging from -3 to +3, but the visualization enhances clarity in both design and pitch contexts. By presenting the results in a structured, visual format, the insights become more accessible, intuitive, and compelling. This not only strengthens the design evaluation process but also makes it easier to communicate findings effectively to potential stakeholders and decision-makers.

MARKING THE HYPOTHESIS VALUES

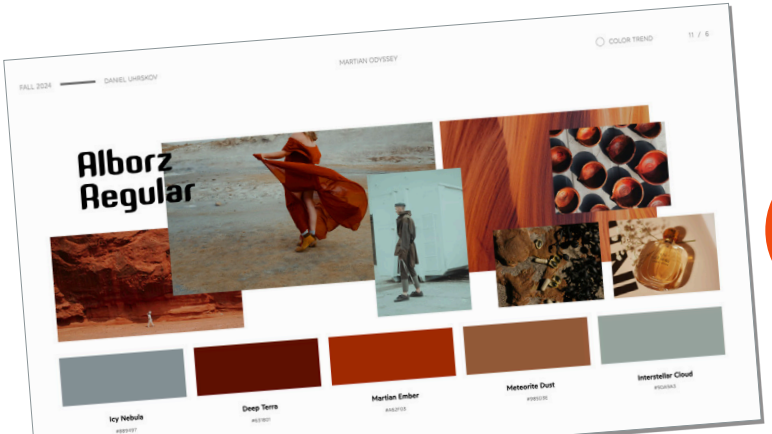
Next, mark the **expected/desired hypothesis values on** the axes based on your desired perception of the visual test medium. Simply place **dots** on the axes to indicate where you believe the test participants will position their responses.

DRAWING THE HYPOTHESIS LINE

Finally, **connect your marked values vertically** across the polarity parameters to **form the hypothesis line**. This visualization clarifies the intended perception path through the parameters, making the test intuitive, structured, and easier to interpret. It represents the emotional and cognitive response you aim to achieve with your design.

SUM UP AND PRACTICAL TAKEAWAYS

- **Clearly define expectations:** The hypothesis sheet serves as a visual benchmark for comparison with actual test results.
- **Limit cognitive load**: Keep the number of polarity parameters under 8 to prevent test participants from becoming mentally and visual fatigued.
- Ensure relevance: The adjective pairs should align with the test's purpose.
- Use the hypothesis line effectively: It helps you visually track discrepancies between your intended message and test participant perceptions.



Practical Example

DEFINING AXES & HYPOTHESIS LINE

The axes and hypothesis line have been carefully set up to measure the perception of the example moodboard (Martian Odyssey). The selected polarity parameters align with the intended design vision, ensuring a structured and meaningful evaluation. By plotting the expected response value (*to the left*) and visualizing them through the vertical hypothesis line (*to the right*), the test gains clarity, making it easier to compare intended and actual perceptions afterwards.

TRANSFER YOUR PAIRS OF WORDS AND PLOT IN THE DESIRED VALUES

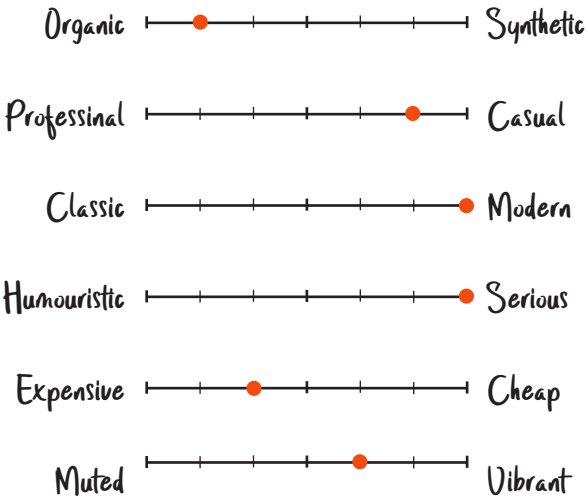


Figure 6: Pairs of words on axis with dots

CONNECT THE DOTS VERTICALLY, AND THE HYPOTHESIS LINE WILL APPEAR

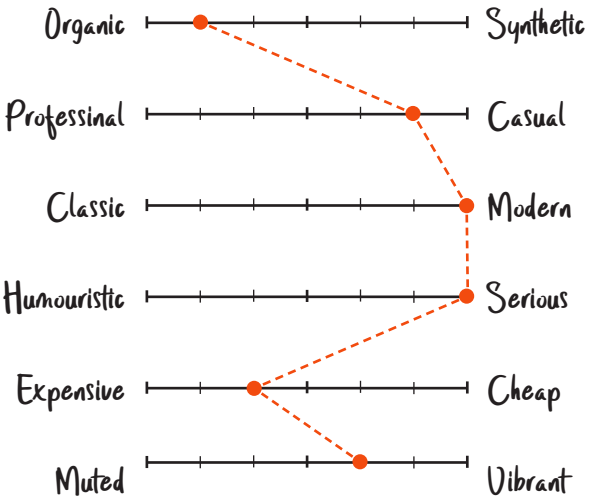
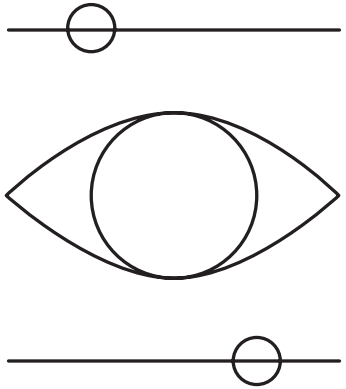


Figure 7: Visual hypothesis line

Step 5

How do you set up the test?

“How do you conduct it, and how many tests do you need to execute?”



Preparing and Conducting the Test

PREPARATION

TEST SHEETS

The first step is to **transfer your selected adjective pairs** onto the blank test sheets (provided at the end of this manual). The axes must be identical to those on the hypothesis sheet but without pre-marked values.

SETTING UP THE TEST ENVIRONMENT

Proper preparation is essential for obtaining reliable results. Ensure that test participants are seated in a quiet, distraction-free space to minimize external influences on their responses. The testing environment should be neutral, well-lit, and free from interruptions to foster focus and authenticity in participant reactions. The facilitator should verify that all materials - whether physical or digital - are properly set up and functioning before the session begins. Additionally, consistent conditions across all sessions help maintain comparability in results. This phase closely aligns with the principles of Think-Aloud testing, where participants are encouraged to articulate their thought processes during the evaluation (34).

INTRODUCING THE TEST TO PARTICIPANTS

Before presenting the test medium, explain the structure of the test using these key points:

1. "You will see a set of visual materials. I am interested in your initial feelings, associations, or thoughts about them. Please respond as spontaneously and transparently as possible. What is your first reaction or emotion?"

2. "You will have to rate certain attributes using axes with 2 opposing words. Mark a dot where you feel your response aligns with the scale in relation to what you see."

3. "Do not overthink it—what was your first impression?"

4. "If you feel that an important value or emotion is missing, use the blank axes to add it and explain it to me in words."

5. "If you have any interesting thoughts that you cannot immediately express, please feel free to verbalize them as you go along."

CONDUCTING THE TEST

Present the visual test medium to the participant. In the VERT test, participants indicate their responses by placing dots on the seven-point scale of each axis. These dots represent their perception of the visual test medium along the selected polarity parameters. Use a marker or pen in a clear, visible color to ensure readability.

If needed, guide them in marking their first response on the test sheet by making dots. Ensure they place the dots on one of the seven measurement points on each axis. Offer clarification if needed, but avoid influencing their responses.

The facilitator should maintain an observational role, ensuring that both verbal and non-verbal cues do not unintentionally influence the participant's responses. Excessive talking, leading questions, or suggestive remarks can introduce bias, compromising the objectivity and reliability of the test results. As Jakob Nielsen aptly puts it: "Shut up and let the users do the talking" (35).

OPEN FORMAT

However, according to the "*Shup up and let the users do the talking*" perspective, VERT is designed to encourage dialogue rather than restrict it. Unlike traditional usability testing,

VERT serves as a catalyst for verbalization, helping test participants articulate their visual perceptions and associations more freely. The structured framework of adjective axes nudges participants toward expressing their intuitive reactions, facilitating richer discussions and deeper insights. These spontaneous moments of reflection and conversation can provide valuable qualitative data, enhancing the overall understanding of how visual elements are perceived.

With reference to the earlier chapter, "*Criticisms and Further Development*" it is essential to reserve space for open-ended dialogue and allow room for oral explanations during the test session. VERT is not merely a survey tool designed for passive completion; rather, it is an experience-oriented method intended to facilitate nuanced discussions and encourage complex feedback. The structured framework provides a starting point, but the true value of the test lies in its ability to open up conversations, making implicit perceptions explicit and fostering deeper engagement between the facilitator and test participant.

ENSURING OBJECTIVITY

Maintaining the integrity of the test requires a neutral and consistent approach throughout the process. Facilitators should avoid leading questions, suggestive tones, or facial expressions that could influence participants' responses. Instructions must be clear and uniform for all test subjects to ensure comparability across results. If a participant struggles to articulate their thoughts, gentle prompting may be used; however, it is essential to avoid suggesting answers or guiding their response, allowing for unbiased and authentic feedback.

DOCUMENTING OBSERVATIONS

Throughout the test, it is beneficial to take detailed notes in a notebook to capture participant feedback, notable reactions, and contextual insights. These notes can provide additional layers of interpretation when analyzing the test results. Especially in relation to the openness just described.

To ensure comprehensive documentation:

- Write down any visible hesitation, confusion, or strong reactions to specific stimuli.
- Note the conversatins and the moments when a test participant share oral feedback.
- Take photos of the test setup, if applicable, to maintain a record of the testing conditions.
- Attach these documentations to the final test sheets at the end of the test process.

PRINT OR DIGITAL EXECUTION?

Based on experiences from the development of VERT, executing the test with printed test sheets is recommended. This approach creates a tangible and structured testing experience, enhancing understanding of the process.

However, digital execution is also possible if the polarity parameters can be displayed and interacted with effectively on a computer. If conducting the test digitally, ensure that:

- The interface is user-friendly and does not interfere with spontaneous responses.
- The axes remain clearly visible and function identically to the printed format.

The visual test medium (Step 1) can be printed or displayed digitally, depending on the format and intended final use. When selecting the presentation method, consider how it aligns with the medium in which the design will ultimately appear to ensure accurate and relevant feedback.

COLLECTING AND VERIFYING TEST DATA

Once the test is completed, all test sheets must be collected and stored for further analysis. Review each sheet to ensure:

- The axes are correctly filled out and no responses are missing.
- Demographic data (if required) has been recorded accurately.
- Any supplementary notes taken during the session are attached to the corresponding test sheet. If inconsistencies or unclear markings appear, facilitators may, immediately after the test, ask the participant to clarify their response.

SUM-UP AND PRACTICAL TAKEAWAYS

- Preparation is essential: Ensure a neutral test environment and verify that all materials are ready before starting.
- Facilitation must be neutral: Avoid biasing the participant—observe rather than guide. Encourage spontaneous reactions: The test is most effective when responses are instinctive and unfiltered.
- Pay attention to the oral feedback from the test participants and note these down. These are as important as the test answers on the sheet.
- Documentation: Use detailed notes to capture observations, participant comments, and testing conditions. Verify data completeness: Check that all test sheets are correctly filled out and that demographic data is included.
- Choose the right execution format: Print is recommended, but digital formats can be adapted if necessary.

WHAT IS THE OPTIMAL NUMBER OF TESTS?

Conducting the appropriate number of tests and with different participants for VERT requires balancing qualitative insights with quantitative validation. Drawing from my observations and analyses of VERT's application, I offer the following recommendations on sample sizes and methodologies.

SMALL SAMPLE SIZES: QUALITATIVE APPROACHES

When using VERT in a qualitative research setting, smaller sample sizes are often sufficient, particularly when the goal is to gain deep insights into user perceptions. Techniques such as Think-Aloud protocols, interviews, and focus groups follow similar methodological principles, as they emphasize interpretative depth over statistical generalization (34).

THE 'MAGIC NUMBER' 5 IN USABILITY TESTING

Jakob Nielsen's research suggests that testing with five participants can uncover approximately 80% of usability issues, as patterns in responses tend to emerge early. This principle, though

primarily applied to usability testing, is relevant for assessing visual perception and branding through VERT. After five to six test participants, diminishing returns set in, meaning that the same feedback is often repeated, with fewer new insights gained. This aligns with VERT's design to detect perceptual patterns and emotional reactions rather than function as a large-scale statistical survey tool (36).

Research on Think-Aloud usability testing indicates that testing with 5 to 10 participants is often enough to detect the majority of patterns in user feedback, as themes tend to saturate within this range (36). Similarly, studies in qualitative interviews suggest that 9 to 15 participants are generally sufficient for reaching thematic saturation, depending on participant diversity and the complexity of the research topic (37). In focus group research, two to three groups typically capture 80% of key themes, particularly when using a semi-structured format (38).

Thus, when VERT is used qualitatively, a small sample of 5-10 participants can provide rich, exploratory insights, especially when combined with Think-Aloud elements or follow-up interviews to probe deeper into responses.

LARGE SAMPLE SIZES: QUANTITATIVE APPROACHES

For quantitative research objectives, such as validating design hypotheses across a broader population, larger sample sizes enhance statistical reliability. The appropriate number of participants is determined by factors such as confidence levels, margin of error, and variability in the dataset (39).

Survey-based research, which shares methodological similarities with large-scale VERT testing, suggests that a minimum of 15-30 participants is required to ensure generalizable results in a medium-scale design evaluation (40). This aligns with research on questionnaire-based user experience studies, where 30 respondents are commonly used as a threshold for ensuring that statistical patterns emerge (41).

However, the number of participants is not the only factor determining validity - researchers must also consider participant diversity, testing conditions, and the consistency of responses (40).

When using VERT quantitatively, a larger sample of 15-30 participants is ideal for confirming perceptual patterns, making the findings more statistically reliable and suitable for broader design validation. Though, you can include many more to further confirm or reject your hypothesis.

ITERATIVE TESTING OVER TIME

Rather than conducting a single round of testing early in the process, it is recommended to integrate iterative testing at multiple stages, allowing for continuous refinement and validation throughout the design process:

Early Concept Testing: Start with 5-10 participants to gather qualitative insights and refine the design.

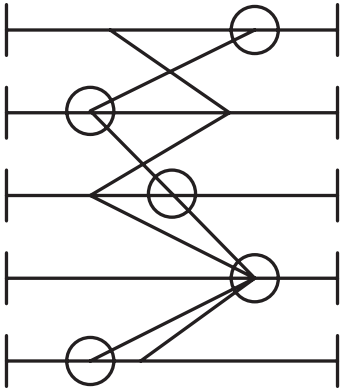
Validation Phase: Follow up with 15-30 participants to confirm findings and ensure statistical reliability before finalizing the design.

For most design-related testing scenarios: 10-20 participants is an optimal range. If the goal is deep insights and exploratory feedback, 5-10 participants are sufficient. If the goal is statistical validation and pattern recognition, 15-30 participants provide stronger reliability.

Step 6

How do you analyze the result?

Whats next?



Visual Evaluation and Further Process

EVALUATION ON THE RESULTS

After completing the VERT test sessions, the final step is to evaluate the collected results and compare them with the visual hypothesis. This phase is critical for understanding how the intended design perception aligns with actual user responses and identifying potential deviations, patterns, or areas for improvement.

VISUALIZING TEST RESULTS: TRANSLATING DATA INTO A VISUAL REPRESENTATION

The strength of the VERT test lies not only in its ability to gather structured responses but also in how effectively the results can be translated into **a visual format**. A purely textual or numerical analysis may fail to communicate key findings in an intuitive and persuasive way, especially to stakeholders involved in the design process. By transforming test results into a clear visual representation, designers can make insights more accessible, compelling, and actionable. The most important aspect of the analysis is always to examine the actual responses and how they position along the axes compared to the hypothesis values. The vertical lines serve as a visual aid, making the test results more intuitive and easier to interpret, but the core evaluation should focus on the data itself and its alignment with the intended perception.

The first step is to plot in the responses from the **Test Sheets** on the **Hypothesis + Evalusation Sheet**. There are two primary methods for plotting responses:

- **Heatmap Representation** – plotting all responses individually.
- **Averaged Line Representation** – calculating the mean and drawing a single representative line.

Each method offers distinct advantages and drawbacks, depending on the test size, purpose, and intended audience.

MANUAL VS. DIGITAL DRAWING OF TEST RESULTS

The choice between hand-drawing or digital visualization of test results depends on the type of study and the number of test participants. Plotting in the answers manually with a pen aligns with VERT's analog methodology, allowing for a direct, flexible, and immediate representation of findings. Digital visualization can offer greater precision and efficiency. Using software tools to plot the responses ensures accuracy in calculations and allows for easier adjustments, comparisons, and professional presentation of the findings. Ultimately, the decision depends on the test facilitator's preference and the requirements of the analysis.

HEATMAP REPRESENTATION

Plotting all individual responses

For a detailed and nuanced visualization, each test response is plotted individually on the **hypothesis sheet**, preserving variations, extremes, and clusters in the data.

HOW TO DRAW THE HEATMAP REPRESENTATION:

- Use the **Hypothesis + Evalusation Sheet**, which contains the original visual intention (the hypothesis line).
- Transfer all individual responses one by one. Connect the dots for one partcipants answers and repeat till you have several vertical lines.
- Lines might cross, merge and/or dissappear.
- The important thing is, that you can separate the hypothesis line from the answers. Use different colors or pencils.

KEY CONSIDERATIONS:

This method is ideal for smaller studies, where individual differences matter. It helps identify clusters, outliers, and overall distribution trends. If responses vary significantly, the heatmap provides a more accurate representation than an

average line. On the other hand, it can be difficult to extract a clear evaluation, since the many lines can be confusing and 'chaotic'.

AVERAGED LINE REPRESENTATION

Calculating the mean response

For a simplified and aggregated visualization, the average of all responses is calculated and plotted as a single representative line.

HOW TO DRAW THE AVERAGED LINE REPRESENTATION:

- Note all responses with numerical values using the VERT scale (-3 to +3 system).
- Calculate the average value for each polarity parameter axis across all test participants. To assist with this calculation, an Excel/CSV template can be downloaded from VERTs website.
- Plot the calculated average value on each polarity parameter on the Hypothesis + Evalusation Sheet and connect the dots with a vertical line.
- It is important that the graphic representation of the average line differs from the hypothesis line.

KEY CONSIDERATIONS

This method is best suited for larger studies, where an overall trend is needed rather than individual variations. It simplifies the analysis but may overlook extreme responses or differing perceptions within smaller subgroups. Though, when you include a lot of answers, the extremes becomes less important because the result i validated with a stronger foundation.

HANDLING PARTICIPANT-ADDED AXES IN THE ANALYSIS

In VERT, test participants have the option to add their own adjective pairs on blank lines if they feel an important aspect is missing from the predefined parameters. While these custom axes provide valuable qualitative insights, they cannot be directly included in the structured numerical analysis, as they vary between participants and are not consistently measured across the test group.

However, these participant-defined axes can still be documented and attached as supplementary qualitative data. They may highlight recurring themes, unexpected perceptions, or areas of interest that were not initially considered in the test design. These additions should be noted in the test observations, potentially influencing future iterations of the test or guiding follow-up discussions in the design process.

This part of the VERT is an important feature, as the axis-system is a catalyst for making test participants share and talk more. Include notes and transcripts in the attachtments on the evaluations sheets.

EXTRACTING KEY INSIGHTS FROM THE COMPARISON OF THE VERTICAL LINES

Remember that the visual hypothesis defines the intended emotional and aesthetic responses of the design. By comparing participant responses (through the **heat map** or **the average line**) with this hypothesis, it is possible to determine:

- **Alignment**: *Do the majority of responses match the expected perception?*
- **Deviation:** *Where do responses differ significantly from the hypothesis?*

PARALLELISM OR INCONSISTENCIES

To refine the design further, the response line(s) should be systematically compared to the hypothesis line, assessing how closely they align or diverge. The parallelism, shifts, or inconsistencies between these lines highlight areas where the intended perception has not been fully achieved. Identifying these strategic points for intervention allows designers to focus their efforts on refining specific visual elements, such as color, typography, composition, or

branding cues, to steer perception closer to the hypothesis.

For example, if the hypothesis states that a brand identity should be perceived as *trustworthy* and *professional*, but participants rate it closer to *playful* and *informal*, this indicates a mismatch that requires further analysis and a new iteration in the design process. Conversely, if participants' responses closely align with the hypothesis - such as rating a luxury product as exclusive and elegant, just as intended - this confirms that the visual choices effectively communicate the desired brand values and require minimal adjustments.

This process is inherently iterative, requiring adjustments that are tested and re-evaluated to progressively fine-tune the design until the target audience consistently responds with the intended values. By analyzing and addressing these key deviations, designers can systematically enhance the visual identity, ensuring that it effectively communicates the desired message while resonating with its intended audience.

To adjust the design effectively and understand why it may not have been perceived as intended - for instance, why a brand identity was not rated as trustworthy - it can be valuable to incorporate qualitative insights alongside the numerical test results. This can be done through follow-up questions or verbal probes during the test, asking participants to explain why they placed their response in a certain position. Such insights help uncover specific visual elements - such as color choices, typography, or layout - that may have influenced perception. By integrating these qualitative observations into the analysis, designers gain a even deeper understanding of how to refine the visual communication to better align with the intended brand values.

TAKING ACTION BASED ON INSIGHTS

Based on the evaluation, the next step is to determine how the findings should influence the design process. You can face these 3 scenarious:

- If the results **align** with the hypothesis, the design can move forward with 'more confidence'.
- If **deviations** appear, adjustments to color, typography, composition, or branding elements may be necessary.
- If the responses are **highly varied**, further iterative testing may be required to refine the design.

SUM UP AND PRACTICAL TAKEAWAYS

- Compare the collected responses to the hypothesis to assess whether the intended visual perception aligns with participant feedback. The response lines should be systematically analyzed against the hypothesis line to identify alignment, deviations, or inconsistencies.
- Choosing the right visualization method is essential. The Heatmap Representation is ideal for detecting outliers and response distribution in smaller studies, while the Averaged Line Representation provides a clear summary of general trends in larger datasets.
- Custom participant-added axes offer valuable qualitative insights but cannot be directly included in the structured numerical analysis. These inputs should be documented separately to provide additional context for refining the design.
- The insights gained should guide design adjustments, focusing on elements such as color, typography, composition, and branding. Identified deviations help pinpoint areas that require refinement to better align with the intended perception.
- Finally, an iterative testing approach is recommended. If the results strongly align with the hypothesis, the design can move forward with confidence. If discrepancies emerge, further adjustments and qualitative follow-ups may be necessary to ensure the design meets user expectations.

PLOTTING IN RESPONSES

Practical Example

These figures builds on previous example, testing the Martisan Odyssey moodboard, demonstrating two methods for plotting VERT results based on responses from 10 test sheets / 10 test participants. On the left, the Heatmap Representation shows all individual responses. On the right, the Averaged Line Representation calculates the mean response and plots a single representative line, offering a clearer overview of general trends while smoothing out individual differences.

HEATMAP REPRESENTATION PLOTTING

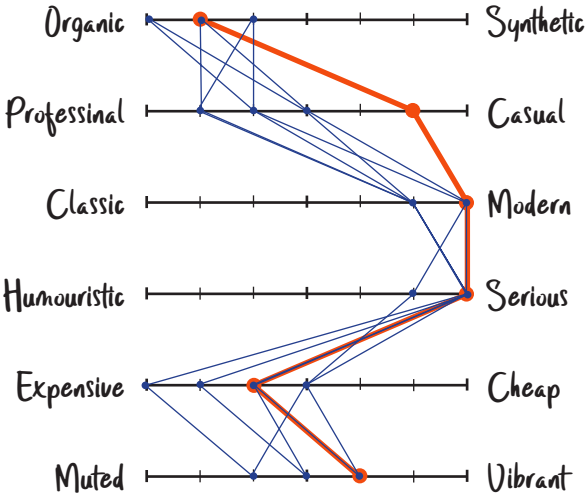


Figure 8: An evaluation with the heatmap representation method

AVERAGE LINE PLOTTING

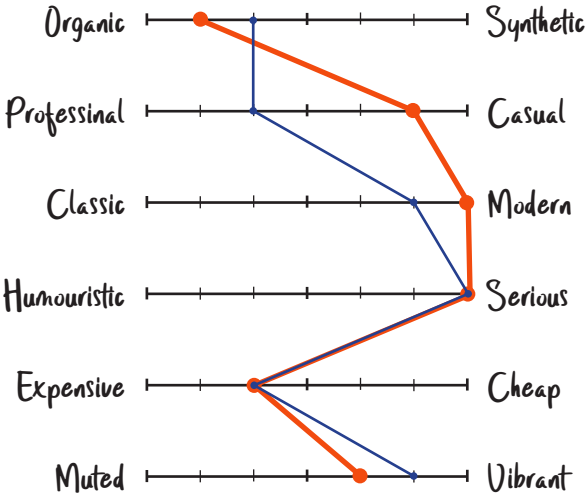


Figure 9: An evaluation with the average line method

KEY INSIGHTS FROM THE VISUAL COMPARISON

The two plotted representations provide different insights into how the test results align with the hypothesis values. Looking at the highlighted areas in the figure (yellow graphics), we can observe key takeaways from the comparison.

The first observation, visible in both the Heatmap and the Averaged Line representation, is that the responses on six polarity parameters align well with the hypothesis line, indicating that the design successfully communicates its intended perception: The test response lines and the hypothesis line run roughly parallel across most parameters. However, there is a notable deviation on the "Professional vs. Casual" axis, where test participants perceive the design as more casual than professional. This misalignment suggests that a design iteration may be needed to enhance the professional appearance, depending on the intended message.

The second observation, which is particularly evident in the Heatmap Representation, is the spread of responses on the "Expensive vs. Cheap" axis. The presence of widely dispersed dots indicates varying interpretations among test participants, meaning that the visual elements used in the design may be ambiguous or context-dependent. This is an important insight for a designer, as it highlights an area where perception is not consistent, suggesting the need for further refinement or audience segmentation in testing.

The Heatmap version is especially useful for detecting response distribution and extreme outliers, while the Averaged Line version provides a cleaner, overall summary of how well the test aligns with the hypothesis. By comparing both methods, designers can make data-driven decisions about which aspects of a design may require further iteration.

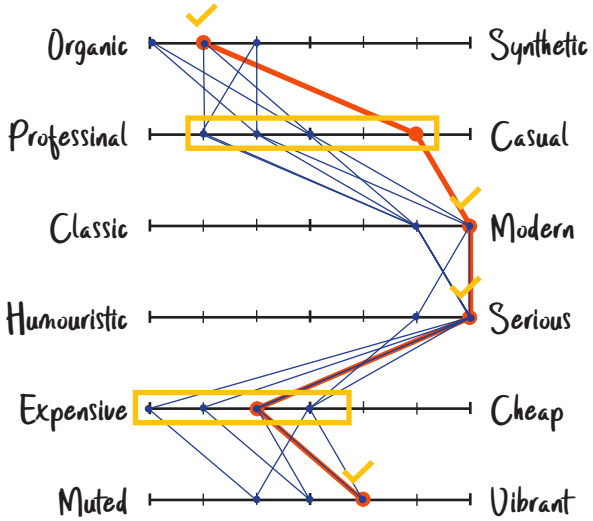


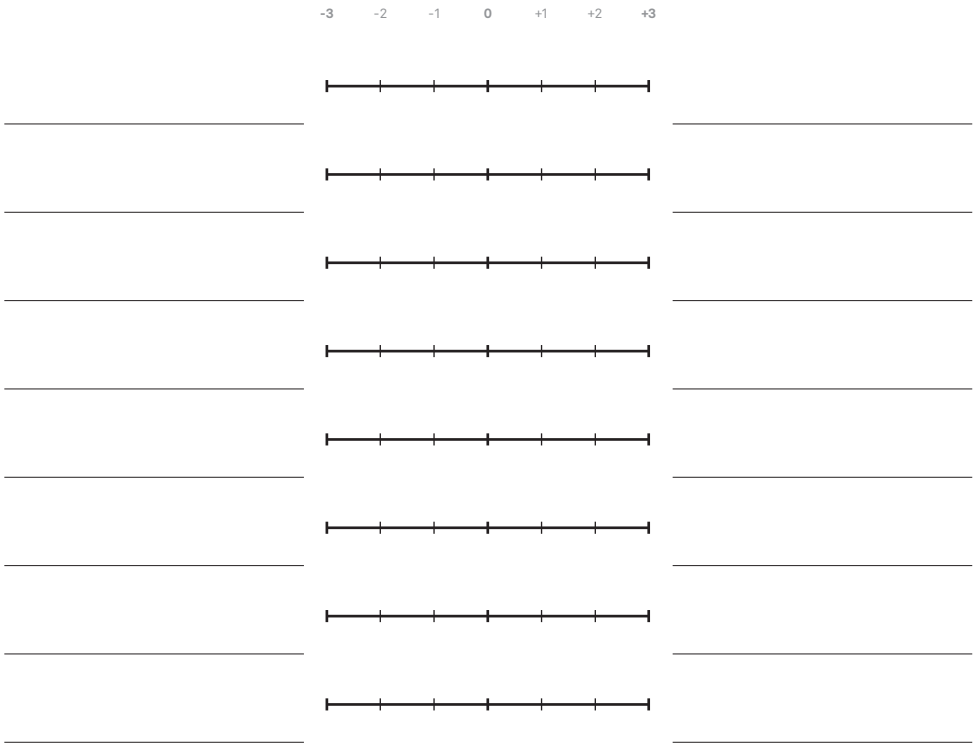
Figure 10: Evaluation with marks indicating key insights

HYPOTHESIS + EVALUATION

TEMPLATE SHEET

HYPOTHESIS + EVALUATION SHEET

- Use this sheet for the **hypothesis** and the **evaluation**.
- **Dot** the intended values on the polarity paramters and conncect them vertically with a **colored line**.
- When the test answers has been collected you can **transfer the result** to compare. Either as an average - or as all indivudal lines.



VERT

A test kit to perform
tests on visual design

Download at www.visualcolorproject.com

Last update: March 2025

TEST

TEMPLATE SHEET

TEST SHEET

Place **dots** according to your experience on the axis.

Add **your own** axis if think, your experience/emotion is not represented.

INFORMATION

- Test date:
- Test location:
- Test medium name:
- Test medium device:

DEMOGRAPHIC NOTIONS

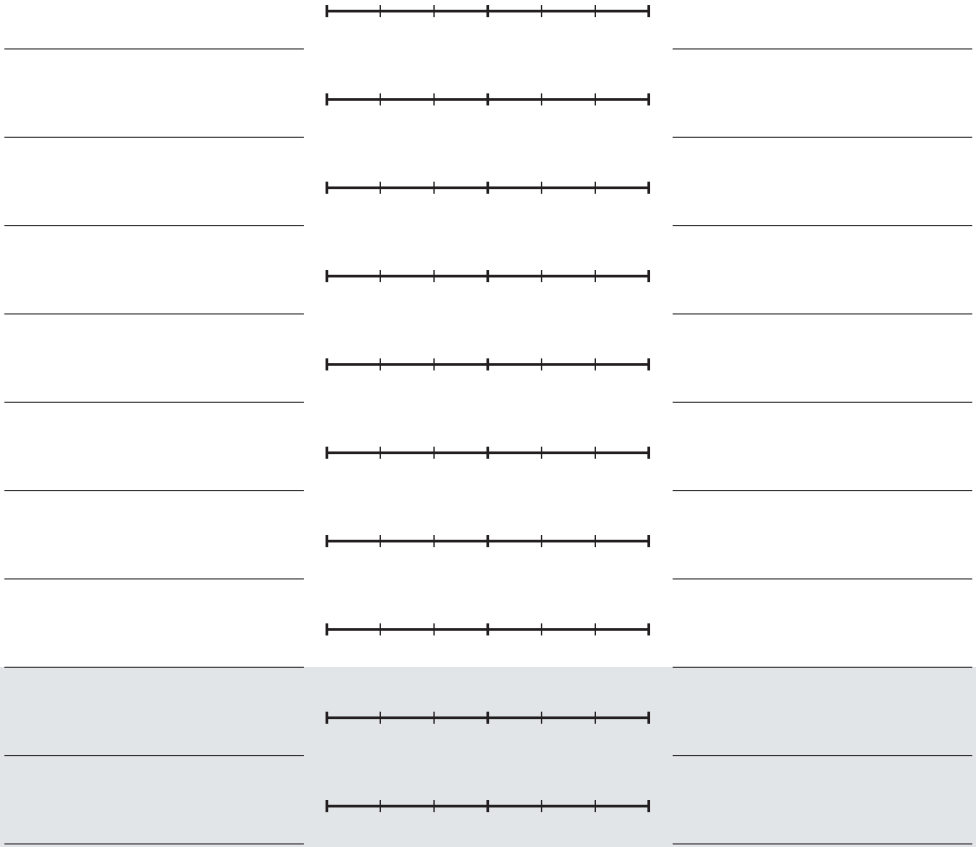
- Age:
- Gender:
- Nationality:
- Occupation:
- Other:

VERT

A test kit to perform tests on visual design

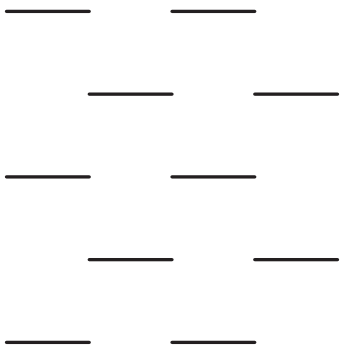
Download at www.visualcolorproject.com

Last update: March 2025



Leave blank for participants additions

Inspiration / example



Adjective Pool

COLOURS

Dark - Light

Sad - Happy

Subdued - Lively

Quiet - Noisy

Serious - Funny

Traditional - Modern

Gloomy - Light

Elegant - Crazy

Violent - Peaceful

Complex - Simple

Minimalist - Sumptuous

Classic - Experimental

Hard - Soft

Retro - Futuristic

Predictable - Surprising

Natural - Artificial

Raw - Processed

Insecure - Confident

Usable - Decorative

Organic - Geometric

Complicated - Simple

Sad - Encouraging

Traditional - Radical

Minimalist - Maximalistic

Elegant - Funky

Gentle - Sharp

Neutral - Colourful

Stable - Alternating

Controlled - Wild

Creative - Conventional

Cosy - Cold

Seductive - Deterrent

Nostalgic - Innovative

Serious - Casual

Concrete - Abstract

Joyful - Gloomy

Comfortable - Uncomfortable

Lively - Lifeless

Boring - Interesting

Dry - Moist

Exclusive - Regular

Relaxed - Formal

Valuable - Worthless

Deep - Superficial

Raw - Refined

Impersonal - Personalised

Bitter - Sweet

Calming - Nervous

Organic - Artificial

Fragile - Robust

Clear - Muddy

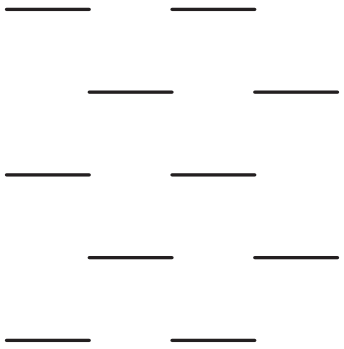
Spicy - Mild

Discreet - Pronounced

Cosmic - Earthly

Hand-drawn - Mechanical

Inspiration /
example



Adjective Pool

TYPOGRAPHY

Serious – Playful

Classic – Modern

Traditional – Innovative

Readable – Decorative

Formal – Casual

Structured – Fluid

Clean – Ornate

Bold – Subtle

Sharp – Rounded

Neutral – Expressive

Static – Dynamic

Symmetrical – Asymmetrical

Balanced – Unbalanced

Elegant – Funky

Restrained – Extravagant

Understated – Pronounced

Sophisticated – Raw

Geometric – Organic

Consistent – Erratic

Dense – Open

Friendly – Serious

Warm – Cool

Approachable – Aloof

Simple – Complex

Delicate – Sturdy

Refined – Rough

Tight – Loose

Light – Heavy

Timeless – Trendy

Familiar – Unusual

Inviting – Intimidating

Casual – Professional

Stylish – Practical

Abstract – Literal

Elegant – Quirky

Rigid – Flexible

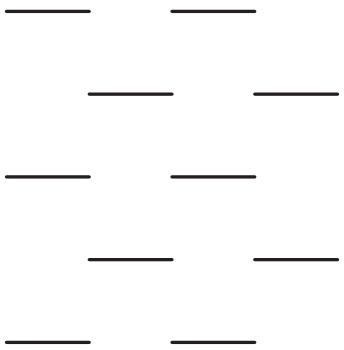
Unassuming – Bold

Mechanical – Handcrafted

Friendly – Hostile

Harmonious – Clashing

Inspiration /
example



Adjective Pool

WEB DESIGN

Professional – Playful

Trustworthy – Bold

Minimalistic – Detailed

Clear – Experimental

Traditional – Innovative

Serious – Fun

Calm – Energetic

Accessible – Exclusive

Balanced – Chaotic

Modern – Retro

Organized – Unstructured

Intuitive – Complex

Formal – Casual

Focused – Distracting

Neutral – Expressive

Timeless – Trendy

Practical – Artistic

Engaging – Passive

Approachable – Reserved

Structured – Abstract

Light – Heavy

Static – Dynamic

Straightforward – Mysterious

Elegant – Funky

Efficient – Decorative

Inviting – Intimidating

Conventional – Adventurous

Reliable – Creative

User-friendly – Confusing

Welcoming – Distant

Refined – Rough

Serene – Exciting

Transparent – Ambiguous

Compact – Spacious

Smooth – Disruptive

Personalized – Generic

Warm – Cold

Friendly – Aloof

Coherent – Fragmented

Harmonious – Contrasting

SOURCES

INTRODUCTION

1. Güss, C. D. (2018). *What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Psychology*. Frontiers in Psychology.

2. O'Brien, L., & Wilson, S. (2023). *Talking About Thinking Aloud: Perspectives from Interactive Think-Aloud Practitioners*. Journal of Usability Studies, 18(3), 113–132.

3. Neumeier, M. (2006). The Brand Gap: How to Bridge the Distance Between Business Strategy and Design. New Riders.

4. Hartelius, A. M. (2022). Visuel identitet i designprocessen. Samfundslitteratur.

5. Schooler, J. W., & Engstler-Schooler, T. Y. (1990). "Verbal overshadowing of visual memories: Some things are better left unsaid." Cognitive Psychology, 22(1), 36-71.

6. Jackson, F. (1982). "Epiphenomenal Qualia." Philosophical Quarterly, 32(127), 127-136.

7. Kay, P., & Regier, T. (2006). "Language, thought, and color: Recent developments." Trends in Cognitive Sciences, 10(2), 51-54.

8. Fine, A. (2017). Color Theory: A Critical Introduction. Bloomsbury.

9. LeDoux, J. (1998). The Emotional Brain. Weidenfeld & Nicolson.

10. Kringelbach, M. (2007). "Emotioner og følelser i menneskehjernen." In T. W. Jensen & M. Skov (Eds.), Følelser og kognition. Museum Tusculanum.

11. Skov, M. (2007). "Følelser og æstetik." In T. W. Jensen & M. Skov (Eds.), Følelser og kognition. Museum Tusculanum.

12. Turner, Niel (2025). *10 Tips for Unmoderated User Testing*. Medium. Link: https://neilturneruxm.medium.com/10-tips-for-unmoderated-user-testing-8ec9bd32c922. Last accessed: 12.03.2025.

13. Teixeira, Fabricio (2024). *The State of UX in 2025: A Love Letter About Change*. UX Collective. Link: https://uxdesign.cc/the-state-of-ux-in-2025-a-love-letter-about-change-8ab5dbe4acdb. Last accessed: 12.03.2025.

14. Brignull, Harry (2011). *Deception vs. Honesty in UI Design*. A List Apart. Link: https://alistapart.com/article/dark-patterns-deception-vs-honesty-in-ui-design/. Last accessed: 12.03.2025.

15. Kelly, George (1955). *The Psychology of Personal Constructs*. Norton.

16. Russell, J.A. (1980). *A Circumplex Model of Affect*. American Psychological Association.

17. Hilleberg, Daniel (2025), Empirical Foundations and Explorative Development of VERT, Link: XXX. Last accessed: 12.03.2025.

18. Van Teijlingen, E. R., & Hundley, V. (2001). *The Importance of Pilot Studies*. Social Research Update, University of Surrey.

19. Gaver, W., Dunne, T., & Pacenti, E. (1999). *Cultural Probes*. Interactions, 6(1), 21–29.

20. Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. MIT Press.

21. Sørensen, F., Mattsson, J., & Sundbo, J. (2009). *ICE and the Experiment Method*. ICE-Project Working Paper, 2009:01, Roskilde University.

22. IDEO. (2012). *Design Thinking for Educators Toolkit*. IDEO.

STEP 1: VISUAL TEST MEDIUM

23. Hartelius, Anne Mette (2022). *Visuel Identitet Designprocessen*. Samfundslitteratur.

24. Rold, Morten (2024). *Interfacedesign: Fra Idé til Digital Prototype*. Samfundslitteratur.

25. Martensen-Larsen, K. (2019). *Nordic Moods*. New Mags.

26. de La Fressange, I., & Montagut, M. (2018). *Maison: Parisian Chic at Home*. Flammarion.

27. Hardik, Dewra (2023). *Decision Fatigue: The Hidden Enemy of User Experience!* UX Design. Link: https://bootcamp.uxdesign.cc/decision-fatigue-the-hidden-enemy-of-user-experience-f62c061d5156. Last accessed: 12.03.2025.

28. Kohavi, Ron, Longbotham, Roger, Sommerfield, Dan, & Henne, Randal M. (2009). *Controlled Experiments on the Web: Survey and Practical Guide. Data Mining and Knowledge Discovery*, 18(1), 140-181.

STEP 2: TEST PARTICIPANTS

29. Dam, Rikke Friis, & Siang, Teo Yu (2023). *Personas – A Simple Introduction*. Interaction Design Foundation.

30. Stickdorn, Marc, Hormess, Markus Edgar, Lawrence, Adam, & Schneider, Jakob (2018). *This is Service Design Doing: Applying Service Design Thinking in the Real World*. O'Reilly Media.

31. Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value Proposition Design: How to Create Products and Services Customers Want*. Wiley.

32. Kongsholm, Louise Byg (2020). *Fra Vugge til Krukke: Livsfaser og Forbrug*. pej gruppen.

STEP 3: FROM WORDS TO THE POLARITY PARAMETRES

33. Guvi, Meghana (2023). *Brainstorming in Design Thinking: A Key to Ideation*. Link: https://www.guvi.in/blog/brainstorming-in-design-thinking/. Last accessed: 12.03.2025.

STEP 5: PREPARING AND CONDUCTING THE TEST

34. Boren, M. T., & Ramey, J. (2000). *Thinking Aloud: Reconciling Theory and Practice*. IEEE Transactions on Professional Communication, 43(3), 261–278.

35. Nielsen, J. (2012). *Thinking Aloud: The #1 Usability Tool. Nielsen Norman Group*. Link: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/. Last accessed: 12.03.2025.

WHAT IS THE OPTIMAL NUMBER OF TEST?

36. Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). *What Influences Saturation? Estimating Sample Sizes in Focus Group Research*. Qualitative Health Research, 29(10), 1483–1496. Sage Journals. Link: https://doi.org/10.1177/1049732318821692.

37. Tokis, Teo. (2021). *Sample Size for Qualitative Usability Testing: How Many Participants Do You Need?*. Useberry Blog. Link: https://blog.useberry.com/sample-size-for-qualitative-usability-testing.

38. Guest, G., Bunce, A., & Johnson, L. (2006). *How Many Interviews Are Enough? An Experiment with Data Saturation and Variability*. Field Methods, 18(1), 59–82. Sage Journals. Link: https://doi.org/10.1177/1525822X05279903.

39. CloudResearch. (n.d.). *Determining Sample Size: How Many Survey Participants Do You Need?*. Cloud Research. Link: https://www.cloudresearch.com/resources/guides/statistical-significance/determine-sample-size.

40. Taherdoost, H. (2016). *Sampling Methods in Research Methodology: How to Choose a Sampling Technique for Research*. International Journal of Academic Research in Management (IJARM), 5(2), 18–27. SSRN. Link: https://doi.org/10.2139/ssrn.3205035.

41. Qualtrics. (n.d.). *How to Determine Sample Size in Research. Qualtrics* Link: https://www.qualtrics.com/en-gb/experience-management/research/determine-sample-size.

DEVELOPED BY DANIEL UHRSKOV, ZEALAND, 2025