



PUBLISHED: SEPTEMBER 28TH · BY SHREYAS KALYANPAD

The aesthetic-usability effect = people rate a good-looking interface as easier to use, whether or not it is.

The problem is that it does not show up in your research as a bias. It shows up as a good result. The prototype is polished, the scores come back high, and the friction ships with it.

A rating and a behaviour are two different measurements. Beauty moves one of them far harder than the other.



She loved it. The clipboard is the half that has to agree.

1. The effect is thirty years old and nothing has fixed it

The study: In 1995, Masaaki Kurosu and Kaori Kashimura at the Hitachi Design Center showed 252 people 26 versions of an ATM screen. Ratings of beauty tracked how easy people thought each one was more closely than how easy it actually was.

Why it still bites: Nielsen Norman Group reports the same thing in modern sessions. A participant fights the navigation, then rates the site highly because the colours and photography are nice.

Do this: Treat a good rating on a polished prototype as unconfirmed, not as a result.

2. Attractive things really do work better, a little

Here is the part most posts leave out. A meta-analysis published in June in the International Journal of Human-Computer Interaction pooled 31 studies and 18,794 people. Better looking interfaces did improve real performance, by a small to medium amount. The authors found no evidence that attractive design hurts performance.

Do this: Stop arguing that visual polish is decoration. It is worth something. It is just worth less than your test says.

3. The gap between the two numbers is the whole story

A 2023 experiment at the University of Basel gave 281 people one of two versions of the same app. Only the look changed. Nothing about the structure or the controls moved.

Perceived usability jumped: $d = 0.86$, a large effect. Task time showed no significant difference. Neither did task score. Across everyone, beauty and perceived usability correlated at 0.79. Beauty and task score correlated at 0.05. Beauty and task time correlated at 0.00.

The opinion moved almost perfectly with the design. The behaviour barely moved.

4. Your test is probably measuring the opinion

What teams do: End the session with rating questions. A SUS score, a five-point scale, "how easy was that?"

Why it fails: Those are the exact measures beauty inflates hardest.

Do this: Report counted behaviour and reported feeling as two separate columns. Never average them into one number, because one of them is doing all the moving.

5. Never run a preference test on a visual comparison

What teams do: Put two designs side by side and ask which is easier to use.

Why it fails: Nobody can answer that by looking. You are collecting what Kurosu and Kashimura called apparent usability, and calling it usability.

Do this: If you have to compare, make people do the task in both, and count what happened.

6. Do not assume you get the average

The same team flagged ten separate factors that could change how much aesthetics helps, and left them open for future work. The gain is an average across very different studies, so it is not a number you can claim for your redesign.

Do this: Measure it in your product, on your task, or do not put it in the deck.

WHAT YOU MEASURED	WHAT IT TELLS YOU	HOW FAR BEAUTY MOVES IT
Ease-of-use rating	An opinion	A long way ($r = 0.79$)
Task score	Behaviour	Almost nothing ($r = 0.05$)
Time on task	Behaviour	Nothing measurable ($r = 0.00$)
Performance, pooled over 31 studies	Behaviour	A small to medium gain

The three-question audit

1. In your last test, which numbers were ratings and which were counted behaviour? Split the list.
2. Were both designs at the same level of visual finish when you compared them? If not, you compared finish.
3. If you delete every rating from the report, does the conclusion survive?

If it only survives with the ratings in, you tested the paint.

Checked against

Read on 28 September 2026.

- Nielsen Norman Group on the aesthetic-usability effect, including the 1995 Hitachi ATM study of 26 designs and 252 participants
- University of Wuppertal on the new meta-analysis of aesthetics and user performance: 31 studies, 18,794 participants
- Perrig and colleagues, Smartphone app aesthetics influence users experience and performance, Frontiers in Psychology