Catchi

User Engagement and Conversion Rates

A meta-analysis via custom tracking and A/B testing

Table of contents

Are measures of user engagement accurate or relevant?

Catchi sought more certainty about how improving user engagement would affect other aspects of website performance. We approached this carefully by assuming that user engagement didn't correlate with any aspects of performance, and that common metrics weren't accurate, unless proven otherwise.

We spent a lot of time critically analysing the data underlying dozens of formal studies and articles written by industry members. We have decided not to name any brands with references. Instead, we encourage readers to do their own exploration with a critical eye that questions assumptions and biases. Here are our own findings:

- **Most studies failed to define user engagement at all. Most others defined 'user engagement' in terms of attention, distraction, visitor frequency, emotion or even addiction.**
- **Most relied heavily on qualitative measures (e.g. questionnaires), and were criticised for being too subjective, while lacking the scale of quantitative measures.**
- **Many implied links between user engagement and other things like sales, using data from measuring User Experience (UX) or Customer Experience (CX) in a much broader sense.**
- **Many referenced formal studies of user engagement, but most of these were in the context gaming, e-learning and social network industries (incentivised by 'gamification').**

Many of these studies are very valid in their respective contexts. For website optimisation, we need a better understanding of when it's valuable to measure user engagement, and how to do so reliably.

Here, we will refer to 'engagement' as users easily finding information that is relevant to their questions and intended actions, so they can spend more of their limited time absorbing helpful information. This is in contrast to 'negative' engagement, where usability is low and users spend more time trying to find information. As noted in the publications we reviewed, it is already well understood that usability is a common bottleneck for engagement, and should be optimised first. This is simply because, if a site has poor usability, users will tend to leave and visit an alternative website, so most non-usability improvements to a website don't realise most of their potential.

How do designers and analysts define and measure user engagement?

We turned to today's industry of professionals and amateurs to understand how they approach user engagement:

- **Most analysts defined engagement in terms of interacting with webpage content (e.g. clicks and scrolling). Many focused on how many times users visit a website, and how close users got to converting. Others prioritised usability and qualitative studies.**
- **The smallest minority focused on how long users spend on a website. It's widely recognised that pageview metrics are too broad, and that time-based metrics are unreliable (e.g. users can open many tabs and disengage for long periods, but still get 'time' attributed to them).**
- **On the contrary, most designers and UX workers agreed that 'positive' engagement was valuable because it is so closely linked to good usability. Many designers agreed that good usability was valuable but, understandably, struggled with a robust explanation as to why.**
- **For designers, the second most valuable aspect of engagement was customer loyalty. Similarly, many designers struggled to provide logic or evidence that directly connected measures of engagement to things like lifetime value, market share, or online reviews.**

The value of things like usability is all but self-evident. However, understanding how valuable engagement could be in various contexts has remained illusive. This begs the question: *"In what circumstances will engagement be a bottleneck for optimising conversions, and how can it be improved with certainty?"*

Scalable click-based metrics are correlated with conversion rates all the time. Click-based metrics are also widely used to estimate relative changes in engagement. Heatmaps are also an extremely common tool for measuring on-page activity and some aspects of engagement with content.

However, it is widely known that these quantitative approaches are often limited by the interpretation of individuals. Most common heatmapping tools don't allow for proper segmentation of data. For example: *"How might a heatmap look different if we excluded all users who have converted in the last 3 months?"*

Especially as mobile use becomes increasingly dominant, most common heatmapping tools also struggle to visualise interactions with content inside dropdowns, popups, expanding forms, dynamic pages, and other forms of 'hidden' content.

How did we define, measure and test user engagement?

To reflect our definition of engagement, we measured it during a number of A/B tests on pages with good usability, and other pages that we improved the usability of. We focused on measuring the amount of time users spent ***actively*** engaging with specific content, on specific pages. We focused on excluding (1) brief periods of time that users spent navigating to pages with the information they were likely looking for, (2) any time users spent with a page open but were distracted away from their device, and (3) any time users spent with a page open but were using a different tab, window or application. We did this by using Google Tag Manager (GTM) to insert a custom script. Here's a simplified version of the code's logic:

- **If a user interacts with their device in any relevant way (e.g. scrolling, moving their mouse, touching their screen, typing information into a form field, etc), start a timer.**
- **If a user stops interacting with their device for a number of seconds, or leaves the current tab in their browser, pause the timer.**
- **While a user is interacting with their device, identify which content is currently displayed on their screen and attribute the time counted to that content.**

By pushing this data via Google Tag Manager (i.e. the data-layer) to Google Analytics (GA), we were able to create relevant goals for A/B tests and segment the results however we wanted. We tested and 'calibrated' our custom engagement tracking to ensure its accuracy on all the pages we tested. We also measured a number of default metrics in GA, and a number of other custom metrics (for brevity, some of these have been omitted from this report).

We created a correlation matrix of all these metrics, using the data captured during several A/B tests. In this way, we could better understand how the correlation between any two metrics changed (or stayed the same) in different circumstances.

In total, we ran four tests on a website owned by a company in the financial services industry. Test #1 was on a short product page, test #2 was on a long product page, test #3 was across all their blog pages, and test #4 was on their homepage. We have also used our custom engagement tracking across other sites, and have re-ran a very similar analysis in the second-to-last section of this white paper. This is simply to provide some insight as to how results can change across different websites and KPIs.

How to interpret correlations vs percentage changes in this report

Changes in correlation values and changes in metric percentages mean very different things. Understanding the differences between them is required to correctly interpret results of various tests.

Changes in metrics are quoted as percentages. The lowest percentage possible is -100%, meaning that a metric has dropped to zero. A percentage of 0% means there was no difference in a test between design A and design B. +100% means that the metric doubled, although larger increases are possible.

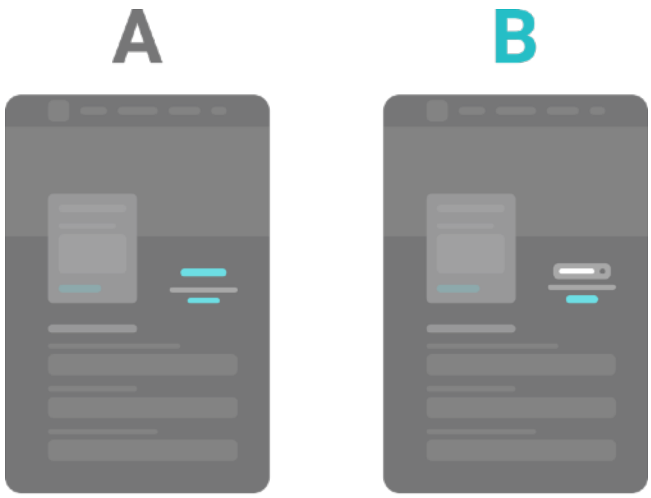
Correlations are quoted as values between -1.00 and +1.00. They describe how one metric changes in relation to another metric, on average. A correlation of -1.00 means there is a perfectly negative relationship between two metrics (i.e. when one goes up by some amount, the other goes down by an equivalent amount). A correlation of +0.50 would mean there is a moderately positive relationship between two metrics (i.e. although they move in the same direction, there is some difference between them that isn't accounted for).

Example: If pageviews and conversions had a correlation of +0.15 then, on average, a +100% increase in pageviews would likely result in a +15% increase in conversions. However, if pageviews decreased by -100% and conversions decreased by -15%, this wouldn't affect the positive correlation of +0.15, because the metrics are still moving in the same direction.

Correlation ≠ Causation: Correlation values tell us how two metrics move together, on average. They do not tell us which metric causes the other metric to move. This is partly why we have included A/B tests in this study. We can say that, changing from design A to design B caused metrics and correlation values to change, with some known degree of certainty.

Summary of test #1 — Short product page

This test focused on giving users more options to enter the conversion funnel once they had engaged with the relevant information they needed to feel ready. Design B replaced a simple CTA with a dynamic CTA, allowing users to use a dropdown to tell the company what type of investment service they were interested in before moving forward. This test was of particular interest in terms of correlating engagement and conversion metrics as the increase in conversion rate from design A to design B was more than double than any of the other tests.



Test results

99.5% Probability of engagement lift

+14% Active engagement time

-20% Scroll beyond 50%

-58% Bounce rate

+47% Conversion rate

+86% Micro-conversion rate

The previous section on page 4 explains how to correctly interpret the changes in metrics above, versus the correlation values below.

Correlations	Design A Active engagement time	Design B Active engagement time
Time on page	0.01	-0.03
Scroll beyond 50%	0.67	0.76
Bounce rate	0.04	0.38

Correlations	Design A		Design B	
	Conversions	Micro conv.	Conversions	Micro conv.
Active engagement	0.28	0.33	0.29	0.5
Time on page	0.10	0.09	-0.20	-0.14
Scroll beyond 50%	0.18	0.04	0.17	0.31
Bounce rate	-0.29	0.01	0.12	-0.02

Brief interpretation

Users who experienced design B were significantly more likely to convert, due to this test's focus on giving users multiple ways to enter the funnel. Although the page had 5 to 6 folds of content, the +14% increase in engagement was largely driven by engagement with the top fold, where design B differed from design A.

The data could suggest that users saw the dynamic CTA on the right-hand side of design B and were encouraged to consider the content on the left. Then, having engaged more, more users were sufficiently informed to consider interacting with the dynamic CTA options in order to enter the conversion funnel.

Engagement was slightly correlated with conversions, more-so than 50% scroll. The correlation between engagement and micro-conversions (e.g. clicking on the CTA) increased from design A to design B, likely because design B resulted in +61% micro-conversions so there was a relatively high amount of data to provide a more accurate correlation. Overall, engagement was more positively correlated with conversions and micro-conversions than any of the other metrics we analysed.

The default time-on-page metric had no correlation with active engagement, strongly suggesting that time-on-page is an untrustworthy metric in similar use-cases. 50% scroll had a moderate-to-high correlation with our engagement tracking, although this was slightly less so for design B, compared to design A.

The bounce rate for design B was much lower than design A. This is partly because design B contained an additional dropdown that users could interact with before leaving. This dropdown was part of the dynamic CTA that users could interact with without clicking through into the conversion funnel. Correlations with bounce rate are also not very consistent, suggesting that it might not be a metric that should be quoted without context provided from other metrics.

Summary of test #2 — Long product page

This test reduce the scrolling distance for a large list of benefits into a list of dropdowns that communicated the key benefits, while offering more information for users who wanted to expand certain dropdowns. Given the great length of the page, we also introduced a table of contents to help users quickly work out if the page likely contained the information they were looking for.

Test results

99.5% Probability of engagement lift

+34% Active engagement time

+16% Scroll beyond 50%

+24% Bounce rate

+20% Conversion rate

+10% Micro-conversion rate

Correlations	Design A Active engagement time	Design B Active engagement time
Time on page	0.37	0.15
Scroll beyond 50%	0.72	0.76
Bounce rate	-0.56	-0.44

Correlations	Design A		Design B	
	Conversions	Micro conv.	Conversions	Micro conv.
Active engagement	0.51	0.21	0.27	0.25
Time on page	-0.30	0.36	0.00	-0.30
Scroll beyond 50%	0.52	0.49	0.27	0.41
Bounce rate	-0.28	0.30	-0.32	-0.36

Brief interpretation

Users who experienced design B were much more likely to find and engage with various pieces of information much further down the page that were relevant to their specific circumstances. The most popular sections of content in design B were engaged with 6 times more than in design A.

Our custom engagement metric correlated well with basic measures of scrolling activity, and this correlation was slightly stronger in design B which was created to have better usability.

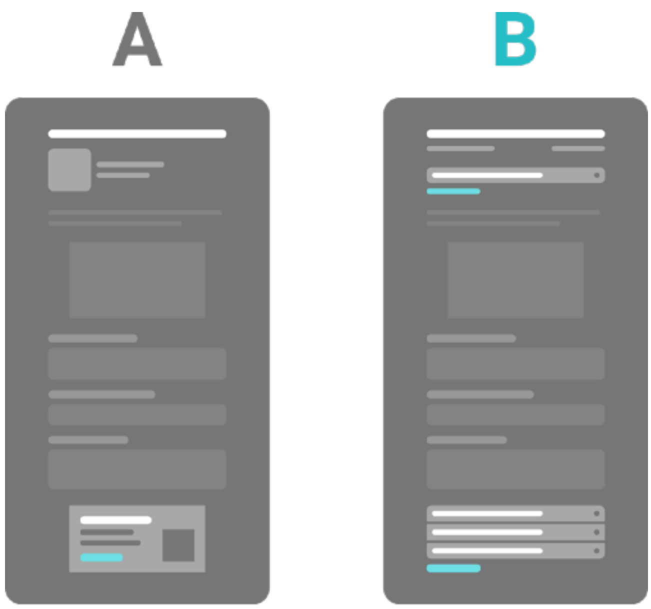
The correlation between engagement and time-on-page was weak in design A and almost non-existent in design B, suggesting that, as the usability of a page improves, basic time-on-page measures become less trustworthy.

There was a moderate level of correlation between engagement and bounce rate for both design A and design B. We expected this correlation to be negative (i.e. inverse) because if users engage more with the content of a page, they are less likely to bounce without interacting with elements on the page (e.g. clicks on dropdowns, links, etc). This is typical, although it is technically possible for both engagement and bounce rate to increase (e.g. users may spend more time scrolling and reading through a page without interacting with on-page elements before leaving).

As with all the tests referenced in this study, we also conducted a much deeper analysis of the results using data from a lot more metrics and tools (e.g. default metrics in GA, custom events via GTM, and heatmaps with custom-triggers).

Summary of test #3 — Blog article pages

This test reduced the height of a tall section of authorship details, and introduced a thin dropdown containing links to other pages that were relevant to the blog article being viewed by users. The bottom of each article also had an image-based section that promoted one of the company's services. This was replaced with a small list of dropdowns containing information that was relevant to both the blog articles and the companies service, along with a CTA for interested users. Design B was created with the aim of making highly-informative content accessible to users, while being as unobtrusive as possible.



Test results

99.5% Probability of engagement lift

+19% Active engagement time

+34% Scroll beyond 50%

-3% Bounce rate

+22% Conversion rate

+13% Micro-conversion rate

Correlations	Design A Active engagement time	Design B Active engagement time
Time on page	0.37	0.37
Scroll beyond 50%	0.84	0.71
Bounce rate	0.00	-0.37

Correlations	Design A		Design B	
	Conversions	Micro conv.	Conversions	Micro conv.
Active engagement	0.35	0.54	0.26	0.71
Time on page	-0.14	0.50	0.45	0.38
Scroll beyond 50%	0.57	0.70	0.36	0.38
Bounce rate	-0.02	-0.11	-0.39	-0.34

Brief interpretation

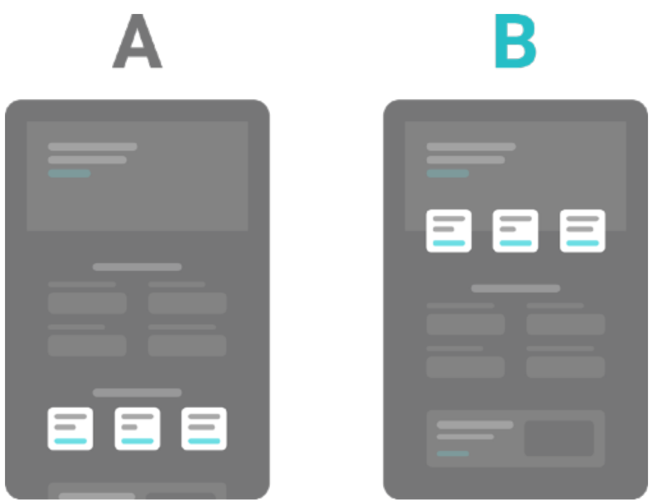
For design B, most users seemed to be able to skim past the authorship information and dropdown at the top of the page, and engage with the article content. Many users who reached the end of the article opened at least one of the of dropdowns at the bottom of the page, and some clicked through. Others scrolled back to the top of the page to navigate elsewhere, and a few of them used the dropdown at the top of the page to look for more information.

Active engagement time was most closely correlated with 50% scroll. Unsurprisingly, this means that users who scrolled at least half-way down a blog page were likely to have higher engagement scores than users who didn't scroll half-way down a blog page. Engagement also had a slight correlation with average time-on-page. Bounce rate had absolutely zero correlation with engagement in design A, and a slight correlation in design B. This is likely because design B introduced a couple new elements that users could interact with before leaving, likely making its correlation more accurate.

In this test, micro-conversions represented click-throughs from the blog articles to the company's main service page. Conversions represented users clicking-through multiple steps to move down the relatively-long funnel. 50% scroll had a moderate correlation with conversions in design A, and a low correlation in design B. This reflects how, although the dropdown at the top of the page received relatively few clicks, those users who did end up interacting with it were likely more interested in converting (rather than using the CTA at the bottom of the page).

Summary of test #4 — Homepage

This test changed the order of sections on the page so that a row of product cards appeared near the fold. The top section of this page contained a lot of white space due to a very large hero image, so this was achieved without pushing the other sections of the page much further down. The aim of the design B was to give new users a better understanding of the company's services within 5 seconds, while providing them with more options to navigate to the section of the site that was more relevant to them. This test was interesting in that the changes between design A and design B had a positive effect on conversions, but zero net effect on engagement time. Instead, users spent the same amount of time engaging further down the funnel.



Test results

0%	Probability of engagement lift
-1%	Active engagement time
0%	Scroll beyond 50%
+19%	Conversion rate
+32%	Micro-conversion rate

Correlations	Design A Active engagement time	Design B Active engagement time
Time on page	0.03	0.26
Scroll beyond 50%	0.62	0.87
Bounce rate	0.09	0.57

Correlations	Design A		Design B	
	Conversions	Micro conv.	Conversions	Micro conv.
Active engagement	0.31	0.51	0.25	0.62
Time on page	-0.12	0.16	0.00	0.08
Scroll beyond 50%	0.31	0.57	0.27	0.65
Bounce rate	0.04	-0.04	-0.10	0.21
Active engagement (excluding homepage)	0.48	0.65	0.34	0.77

Brief interpretation

Design B saw a 19% increase in conversion rate, but only the 32% increase in micro-conversions was statistically significant. Nevertheless, we are primarily interested in the correlations between metrics, more so than their changes from design A to design B. Although there was no change in overall engagement time, design B resulted in a large decrease in engagement on the homepage, balanced by a large increase in engagement on a number of key product pages. This is a key difference in the context of this test, as we expected engagement on the homepage to decrease if users navigate more efficiently to other pages with more relevant information.

The correlation between engagement and conversions is noticeably stronger if the homepage is excluded from the overall engagement data (as shown by the additional grey row in the table above). This supports the hypothesis that active engagement further down a funnel tends to result in higher conversion rates. It also hints at the idea that users have a finite amount of time and energy that they have (sometimes subconsciously) allocated towards trying to complete an intended action. However, for consistency in our meta-analysis across all tests (in the following sections of this report), this additional row of greyed-out correlations has not been used.

Engagement and 50% scroll were highly correlated. Both metrics had moderate correlations with conversions (across design A and design B). Other metrics showed little-to-no correlations with engagement or conversions.

What were the limitations of this meta-analysis?

There is a subjective element to this meta-analysis, and so findings depend somewhat on opinion. Specifically, for each of the four tests, design B was derived from small-to-moderate changes to design A, in order to improve conversion rates without negatively impacting usability. For example, to minimise any potential confusion, no unconventional elements were used in any designs.

As such, the underlying assumption of this meta-analysis is that an increase in active engagement time represented a positive user sentiment. Each design was also thoroughly tested across several device/OS/browser combinations to ensure there were no bugs contributing to users getting stuck, and therefore increasing active engagement time.

On average, each test was run for about 4 weeks on a website with about 200K sessions per month. To avoid the incentive of optimising conversion rates more than user experiences, these tests were run on webpages that are relatively high-up in the funnel. Although our focus was on active engagement tracking, three out of the four tests achieved statistically significant increases in conversion rate. Additionally, all four tests achieved statistically significant increases in micro-conversion rates (e.g. clicking-through and moving down a conversion funnel).

Disclaimer — As an optimisation consultancy with expertise in development, analytics and user experience, Catchi will assume that each test was successfully designed to affect both conversion rate and user engagement in a neutral or positive way. The designs and absolute figures, however, have been omitted to respect confidentialities.

Which metrics correlated best across all tests?

We investigated many metrics and, for brevity, some have not been included in this report. The closest correlations across all tests were between conversions and our custom measures of active engagement time. On average, 50% scroll also correlated closely with conversions, however, it was less consistent than active engagement (i.e. 50% scroll had a greater variance across all tests).

Active engagement was a strong predictor of conversions, especially micro-conversions. Because it measured engagement at both the page and the content-section level, it offered highly-granular insights into changes in user behaviour. It is similar to heatmapping but, due to its tight integration with Google Analytics, it also offered full control via filters and segmentation.

As discussed in the previous section, it is reasonable to assume that, for all tests, the usability of each design B was better (or at least similar) to design A. We also analysed usability changes for each test individually, but this analysis is beyond the scope of our investigation.

The correlation between engagement and conversions was relatively stable across design A and design B, for all tests. The correlation between engagement and micro-conversions was much stronger in design B for every test. This discrepancy is partly due to micro-conversions being more common than conversions (which are a step later in the funnel). Tests were also conducted on pages relatively high-up in the funnel, so we expected a greater impact on micro-conversions.

This investigation across a variety of tests warrants further analysis, and makes a good case for active engagement tracking. Accurate and reliable metrics are a valuable asset for optimisation.

	Correlations with active engagement time			
	Design A		Design B	
	Conversions	Micro-conv.	Conversions	Micro conv.
Test #1	0.35	0.54	0.26	0.71
Test #2	0.51	0.21	0.27	0.25
Test #3	0.28	0.33	0.29	0.50
Test #4	0.31	0.51	0.25	0.62
Weighted average	**0.37**	**0.35**	**0.27**	**0.46**
Standard deviation	**0.13**	**0.15**	**0.02**	**0.19**

What about other metrics and websites?

We correlated several default metrics against conversions and engagement, on a vastly different website, with the same custom engagement tracking. This second website offers tangible products (not services) and so its conversions are defined in terms of transactions (not account creations). It is also more global in nature, so it has significantly more traffic and conversion volume. We analysed this second website as a whole, in the absence of A/B testing, over a 2-month time period (about twice as long as the A/B tests explored in this white paper).

Overall, our active engagement time tracking had the strongest correlation with transactions. We chose to compare this result to other metrics that we knew would correlate positively with transactions. For example, because the transaction funnel on this ecommerce website is 5 pages long, sessions with transactions tend to have significantly greater pageviews per session and session durations. Nevertheless, active engagement time had a stronger correlation with transactions than both of these metrics.

! **We generally recommend against comparing the figures below to those in other tables. This table shows how our engagement metric held up on a vastly different website, with different conversion types. Please see page 4 on how to correctly interpret correlations.**

Correlations with transactions (on a different ecommerce website)		
Active engagement time	Pageviews per session	Session duration
0.568	0.488	0.295

All three metrics in the table above correlated with transactions slightly more than revenue. This is likely because a user's willingness to convert, and how much they're willing to spend, are affected by a different set of factors. After all, most users have a budget cap that is relatively inelastic. They could use their budget to purchase a product on one website, or another, but not both. How easy it is for users to find the information they're looking for and convert, likely influences which website they decide to convert on. As such, this likely affects online market share more than average order value or even lifetime value. Catchi has explored the optimisation of lifetime value in a previous white paper, and we're working to find more relationships with transaction value and frequency.

What are the key takeaways?

Continuing to this day, most digital teams and researchers seem to take a high-level view of user engagement. In the context of web optimisation, this view is limited in its applicability because the metrics and analytical methods used don't consider the 'root' of true user engagement. 'Positive' engagement and 'negative' engagement are largely qualitative measures that are dependent on our underlying definition of engagement. This meta-analysis shows that active engagement time can be defined and measured in a quantitative way that is scalable and useful, especially compared to other metrics that are commonly used in the industry (although, for brevity, we have not included all the data from our analysis in this white paper).

Our hypothesis for active user engagement was based on the fact that users have finite money, energy and time. Users usually visit a site to find and absorb information to help them make decisions, and complete one or more actions from a set of considerations. Users' finite resources result in a trade-off, whereby each consecutive action is dependent on previous actions and the outcomes of those actions. Users value their own finite resources more than they value most brands and platforms (although, for a minority of users, their livelihood might be invested in those brand's reputations and platforms).

Like many optimisation specialists, Catchi strives to identify win-win opportunities for both businesses and their users. We have previously published a white paper exploring the potential benefit for businesses to put users' priorities ahead of their own, because of the disproportionate power of lifetime value (resulting from return customers and a cascade of positive externalities). Engagement is highly-valuable currency for users with limited resources. This is especially the case for limited time. Users can always gain more energy and money over time, but that time cannot be given back. By respecting the value of users' time, businesses can make users' experiences more efficient, leading to more informed decisions and completed actions.

If accurately measuring active engagement time is valuable, why is it so uncommon? Businesses themselves are made up of users with finite money, energy and time too. Doing so is an investment, and it's not easy. However, by doing so, you can give users a more efficient experience, while earning longer-lasting returns and a competitive advantage.

Authored by Hadleigh Waldegrave, UX & Conversion Consultant at Catchi