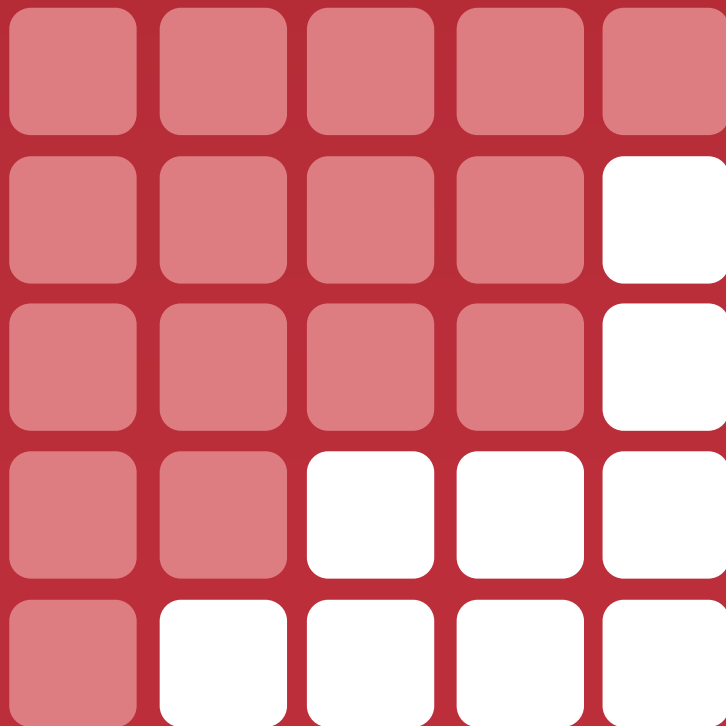


MEASURING THE IMMEASURABLE: VISITOR ENGAGEMENT

ERIC T. PETERSON, WEB ANALYTICS DEMYSTIFIED
JOSEPH CARRABIS, NEXTSTAGE GROUP



Research and Analysis from

WebAnalyticsDemystified

The Web Analytics Thought Leaders

www.webanalyticsdemystified.com

EXECUTIVE SUMMARY

Without a doubt, “engagement” has been one of the hottest buzzwords in digital advertising and marketing in the past 18 months. Forrester Research has written about it, companies founded to measure it, and countless arguments spawned just seeking a reasonable working definition of the term to apply in a meaningful way to the online channel.

Unfortunately, despite the intense level of interest in the subject, few real gains have been made towards developing a practical and useful measure of engagement that can be applied to billions of dollars of advertising, marketing, and technology investments made annually on the Internet. While solutions exist—notably the Evolution Technology™ developed by this document’s co-author Mr. Joseph Carrabis—most are relatively unknown and some are not easily integrated with the widely deployed digital measurement solutions in the marketplace today.

Until now.

In 2007 primary author Eric T. Peterson saw an opportunity to create a tangible measure of engagement that:

- Can be deployed throughout individual organizations;
- Add value to the business’s current understanding of visitor behavior;
- Provide additional evidence for the need to make substantive changes to the web site and other digital marketing efforts

Given the widespread deployment of digital measurement solutions like Omniture, Coremetrics, WebTrends, and Unica, Mr. Peterson believed that this measure of engagement needed to be:

- Practical to calculate using commonly available technology;
- Applicable operationally to the points of leverage currently available to online marketers

This white paper describes exactly that: an *operational measure of micro-engagement* designed to advance the marketers’ knowledge of visitor interaction and support increasingly complex investments into advertising, marketing, content deployment, and technology, especially in situations where traditional measures like “conversion” are unavailable, impractical, or inappropriate. The balance of this document describes Web Analytics Demystified’s measure of Visitor Engagement, including the metrics calculation and use.

The ideal outcome for any reader is to recognize that visitor engagement is a tremendously powerful metric that, when practically applied, can advance an organization’s understanding of their audience and their overall investment in any online channel.

TABLE OF CONTENTS

Executive Summary.....	2
Table of Contents.....	3
Introduction	5
A Note About How this Calculation Can Be Used	6
What Does Engagement Even Mean?.....	7
Why a New Measure of Engagement?.....	10
The “Basic” Measures of Engagement	10
Session Duration.....	11
Page Views per Session	12
Visits per Visitor	12
Conversion Rate.....	13
Customer Satisfaction.....	14
The Visitor Engagement Calculation.....	16
No Way! There is not One Universal Definition of Engagement!.....	16
The Fundamental Assumptions.....	17
The Tools You Need to Make this Calculation	18
The Component Metrics Explained.....	19
The Click Depth Index.....	19
The Duration Index	22
The Recency Index.....	23
The Loyalty Index	24
The Brand Index.....	26
The Feedback Index.....	27
The Interaction Index.....	29
Setting Thresholds for Click-Depth and Duration Indices.....	30
Excluding Individual Indices.....	31
Putting It All Together: Making the Calculation	31
Examples of Visitor Engagement in Action	34

Visitor Engagement and Segmentation Strategies..... 39

About the Authors..... 41

 Joseph Carrabis..... 41

 Eric T. Peterson..... 41

About NextStage Group 41

About Web Analytics Demystified..... 42

Appendix..... 43

 Rule Set..... 44

 First Modification: Standardize Time..... 45

 Second and Third Modifications: Group Like Variables Together, Incorporate Scaling..... 46

 Fourth Modification: Non-Standardizable Variable Forms 49

 Final Forms..... 50

Figures..... 52

INTRODUCTION

In the web analytics and audience measurement community, few issues in 2007 generated more interest and debate than the conversation about measuring visitor engagement. Initiated largely by Microsoft's Robert Scoble in his October 2006 post titled "New audience metric needed: engagement"⁽¹⁾ and followed quickly by Forrester Research's Jeremiah Owyang in his summary of a social media roundtable where the group discussed the question "what should we measure?" in the context of social media and Web 2.0⁽²⁾. Both gentlemen noted that currently available web analytics tools did little to help them measure new media in a valuable way and cited the need for a measurement solution that would capture more than just clicks.

While Scoble and Owyang did a great deal to highlight the need for a new measure of visitor interaction, these respected gentlemen did little to transform engagement into something tangibly measured using web analytics tools. That work was undertaken in part by Eric T. Peterson of Web Analytics Demystified in a widely cited blog post titled "How do you calculate engagement? Part I"⁽³⁾ where he proposed that it is possible to derive a measure of visitor engagement based on the site's specific business objectives using commonly available web analytics tools.

Ironically, NextStage has been making this exact type of measure of engagement since 2001 using their sophisticated tool set and patented Evolution Technology. But the NextStage solution is de-coupled from traditional web analytics and is only now becoming widely known. The company's founder and co-author of this document, Mr. Joseph Carrabis, freely admits little knowledge of web analytics -- a primary driver behind this document and the Web Analytics Demystified/NextStage collaboration.

With support from Mr. Carrabis, Web Analytics Demystified has developed a unique framework for measuring visitor engagement appropriate for use across many different types of measurement systems as well as easily modified to suit any site's specific business objectives and goals. The framework leverages Mr. Peterson's original definition of online engagement first proposed in December 2006:

Engagement is an estimate of the degree and depth of visitor interaction on the site against a clearly defined set of goals.

This definition was made measurable by Web Analytics Demystified and, thanks in part to the work of Mr. Carrabis, refined to the following calculation:

$$\Sigma(C_i + D_i + R_i + L_i + B_i + F_i + I_i)$$

¹ <http://scobleizer.com/2006/10/25/new-audience-metric-needed-engagement/>

² <http://www.web-strategist.com/blog/2006/12/06/factiva-social-media-roundtable-helps-to-answer-what-should-we-measure/>

³ <http://blog.webanalyticsdemystified.com/weblog/2006/12/how-do-you-calculate-engagement-part-i.html>

For the more mathematically inclined:

$$E = \frac{1}{t} \left(\sum_{j=1}^v w_j (1 + 1/\Delta T_j) + \underbrace{\left(\sum_x L_x + \left(\sum_f F_f \right) + \left(\sum_c I_c \right) + \left(\sum_b B_b \right) \right)}_{w_u} + \sum_{k=1}^z \frac{d_k}{d_k} \right)$$

In human terms this calculation says:

"Visitor Engagement is a function of the number of clicks (Ci), the visit duration (Di), the rate at which the visitor returns to the site over time (Ri), their overall loyalty to the site (Li), their measured awareness of the brand (Bi), their willingness to directly contribute feedback (Fi) and the likelihood that they will engage in specific activities on the site designed to increase awareness and create a lasting impression (Ii)."

The components of the Visitor Engagement calculation are:

- **Click Depth Index:** Captures the contribution of page and event views
- **Duration Index:** Captures the contribution of time spent on site
- **Recency Index:** Captures the visitor's "visit velocity"—the rate at which visitors return to the web site over time
- **Brand Index:** Captures the apparent awareness of the visitor of the brand, site, or product(s)
- **Feedback Index:** Captures qualitative information including propensity to solicit additional information or supply direct feedback
- **Interaction Index:** Captures visitor interaction with content or functionality designed to increase level of Attention the visitor is paying to the brand, site, or product(s)
- **Loyalty Index:** Captures the level of long-term interaction the visitor has with the brand, site, or product(s)

Each of these component indices is described in detail in the section on "The Visitor Engagement Calculation."

A NOTE ABOUT HOW THIS CALCULATION CAN BE USED

Early on, Mr. Peterson decided that the general equation and his body of work on the subject of measuring visitor engagement would be available to all, essentially an "open source" project. This decision was made in part because few authors in web analytics have ever suggested a completely new metric or framework for measurement. The net result is a collaborative effort, available to all and patentable by none, which Mr. Peterson hopes will shape the future of measurement in the digital realm.

Any questions about this document or the calculations described herein should be addressed to the authors.

WHAT DOES ENGAGEMENT EVEN MEAN?

Frequently when discussing engagement, authors will refer to the notion of “Customer Engagement” and “Brand Engagement.” Customer engagement is currently defined in the Wikipedia as “the engagement of customers with one another, with a company or a brand. The initiative for engagement can be either consumer- or company-led and the medium of engagement can be on or offline”⁽⁴⁾ and is usually framed in a longer-term context. Multiple definitions for customer engagement have been proposed⁽⁴⁾, including:

“Engagement is turning on a prospect to a brand idea enhanced by the surrounding context.” (Advertising Research Foundation)

which is decidedly vague, and:

“[Engagement is] repeated, satisfied interactions that strengthen the emotional connection a customer has with the brand.” (Ron Shevlin)

which sounds good but is more or less impossible to measure, to Forrester’s own:

“Engagement is the level of involvement, interaction, intimacy, and influence an individual has with a brand over time.” (Brian Haven and Suresh Vittal)

which is by the author’s own admission, more or less impossible to measure due to the fact that even when data is available, it is contained in disparate systems. Most marketing organizations lack the sophistication to combine this data, much less have the ability to take advantage of it.

Furthermore, each of these definitions intermix the notion of “customer” and “brand” as if they are the same. The Wikipedia defines brand engagement as “a term loosely used to describe the process of forming an attachment (emotional and rational) between a person and a brand”⁽⁵⁾ which is measurable albeit not using practical and easily applied technology.

The authors recognize that there is great value associated with measuring and managing customer and/or brand engagement across online and offline channels, working to strengthen the emotional connection a customer has with the brand by influencing their levels of involvement, interaction, intimacy, and influence. But, being practically minded, and given the great difficulty associated with measuring a customer’s engagement, the authors prefer to focus efforts on a more operational, more practical measure.

To this end, we propose that the most useful definition of engagement is that of “attentional engagement”, something that has been well recognized in various social disciplines since the mid 20th century (if not earlier) and has more recently been clarified by Mr. Carrabis. Simply put, someone is engaged when they have focused their attention on something:

⁴ http://en.wikipedia.org/wiki/Customer_engagement

⁵ http://en.wikipedia.org/wiki/Brand_engagement

Engagement is the demonstration of Attention via psychomotor activity that serves to focus an individual's Attention.

This definition of engagement depends on the definition of Attention (similarly found in the literature as):

Attention is a behavior that demonstrates that specific neural activity is taking place.

Unlike the definitions cited above, Mr. Carrabis's definition is extremely precise—by focusing our attention we become engaged. Engagement thus defined can go from internal dialogue to simple hypnotic attention to fugue states. Some real-world examples include⁽⁶⁾:

- You could be engaged with a book, as demonstrated by losing track of time, missing meals (without realizing it), not hearing someone calling your name or tapping your arm (to get your attention);
- You could be engaged with planning your day or reviewing your day while driving such that you lose track of where you are (your non-conscious is handling standard driving functions). Your brain doesn't signal that your conscious attention is needed elsewhere unless some threshold event – such as a police car behind you flashing its lights – is recognized;
- You could be engaged with a web site, demonstrated by various psychomotor behavioral cues indicative that attention is focused on the site.

Mr. Carrabis's definition of engagement then becomes:

Engagement is the demonstration of Attention via psychomotor activity that serves to focus an individual's Attention.

Attention is a behavior that demonstrates that specific neural activity is taking place.

Mr. Peterson originally proposed in 2006 that engagement, relevant to web sites and the field of web analytics, could be defined as:

Engagement is an estimate of the degree and depth of visitor interaction on the site against a clearly defined set of goals.

A more accurate version of this statement, given that "degree" and "depth" are essentially the same in this context, would be:

Engagement is an estimate of the ~~degree and~~ depth of visitor interaction on the site against a clearly defined set of goals.

As presented, the author's definitions are complimentary. Mr. Carrabis has described the most general sense of engagement and Mr. Peterson has applied the definition directly to a specific object: a web site. In this specific case the demonstration of Attention will be measured as a function of the "degree and depth of visitor interaction against a clearly defined set of goals."

⁶ Note that the recurring theme in these examples is that the individual isn't aware they are engaged in an activity. Engagement is non-conscious because the individual's focus of attention is on what they're doing, not the fact that they're engaged with what they're doing. A colloquial expression for engagement might be "single-tasking".

While Mr. Carrabis discusses the relationship between the two definitions extensively in this document's Appendix, we believe this definition of engagement appropriate for the online channel for the following reasons:

- No absolute measures of engagement currently exists within traditional web analytics so a relative estimate is the only practical calculation that can be made;
- "Depth" of visitor interaction can be measured using commonly available quantitative and qualitative research technologies;
- Commonly available technology can be augmented using more robust data sources which are available and able to be integrated into the primary visitor data set;
- Site goals, and by extension, engagement goals differ from site to site, but can be measured on a site-by-site basis

Upon his initial reading of Mr. Peterson's work, Mr. Carrabis pointed out that engagement transcends object and device (for example, a web site being browsed using a notebook computer) so the measurement of engagement must be able to transcend objects and devices as well. Fortunately the mathematics support this idea and thus a more appropriate version of Mr. Peterson's definition is:

Visitor Engagement is an estimate of the depth of visitor interaction ~~on the site~~ against a clearly defined set of goals.

Applied to a single site or group of closely related sites (using a system able to record very granular details about the activities of individual visitors) results in data similar to that seen in Figure 1:

Site Hostnames	Pct Highly Engaged	Visitor VE
blog.webanalyticsdemystified.com	4.5%	12.1%
daniel.webanalyticsdemystified.com	7.4%	15.1%
jobs.webanalyticsdemystified.com	14.1%	17.1%
judah.webanalyticsdemystified.com	2.5%	9.8%
paul.webanalyticsdemystified.com	5.0%	14.6%
thefutureof.webanalyticsdemystified.com	8.4%	15.3%

Figure 1: Visitor Engagement tracked across multiple sites in a domain.

Since the original publication of this work, Mr. Carrabis has revised this calculation to make it mathematically correct as well as to extend the calculation's ability to be both *extendable* and *extensible* for uses not previously imagined by Mr. Peterson. An explanation and the result of Mr. Carrabis's work is found in the Appendix. The net result is that the equation above has been shown to work perfectly well when applied to a single source of information (for example, web server log files, mobile phone transaction logs, etc.).

WHY A NEW MEASURE OF ENGAGEMENT?

One question we have not yet addressed in this document is “why do we need a new measure of engagement?” Why can’t everyone just continue to use simple measures like average time spent and page views per session? In practice, don’t most companies still lack the sophistication to take advantage of even these basic measures?

We realize that the work described in this document really isn’t for everyone. The vast majority of companies *can* and *probably should* continue to rely on the basic engagement measures provided with every good web analytics package while they work to refine their understanding of their site, their business, and the nuances of running a digital business. To this end, we have outlined a handful of “basic” measures of engagement in widespread use in the following section.

However, in our experience working with hundreds of companies at varying stages of maturity in their use of web analytics, both authors are meeting an increasing number of companies that do have the level of sophistication required to take the measurement of engagement to the next logical level. Couple this with an increasing body of evidence that a majority of companies believe that online customer engagement is either essential or important to their organization and you begin to appreciate the need for the measures of engagement described in this document.

Given the rapid change of pace on the Internet and the rapid increase in deployment of technologies that break the Web 1.0 measurement paradigm, now more than ever marketers, advertisers, and business owners need to explore more robust measures of success that can be applied to their online investments. As the use of novel technology continues and the level of sophistication increases, the simple measures that have served well for so long begin to increasingly fail, leaving companies making decisions in a data vacuum.

A great example of the use of newly defined metrics is that of Billy Beane and the Oakland A’s described by Tom Davenport and Jeanne Harris in *Competing on Analytics: The New Science of Winning*. Beane and others improved their competitive abilities by incorporating derived metrics like “on-base percentage” and “on-base plus slugging percentage” into their player assessment program. By leveraging the available data in new and innovative ways, Beane and others (including the New England Patriots) have been able to dramatically improve the capabilities of their respective businesses by several traditional measures.

Put another way, measures of engagement are included in your web analytics package today—session duration, page views per session, visits per visitor, conversion rate, and customer satisfaction. These are the traditional measures, just like “runs batted in” and “batting average”; **the Visitor Engagement calculation is a new measure that has the potential to completely change the way marketing, advertising, and technology deployment is done on the Internet.**

THE “BASIC” MEASURES OF ENGAGEMENT

Five metrics are commonly used as measures of engagement in a variety of situations online. The first three—session duration, page views per session, and visits per visitor—are components of the engagement framework described in this document and are reasonable proxies for engagement as long as you’re aware of their limitations. The latter two—conversion and satisfaction—are often used as proxies for engagement but we believe they are complimentary metrics inappropriate for use in this context. In all cases, these definitions come from traditional web analytics.

SESSION DURATION

Session duration is commonly thought of as the amount of time a visitor spends browsing the web site. Session duration was defined by the Web Analytics Association in their 2007 Web Analytics Definitions document as “the length of time in a session” typically calculated by taking “the timestamp of the last activity in the session minus the timestamp of the first activity of the session” and not reported when “there is only one piece of activity in a session.”

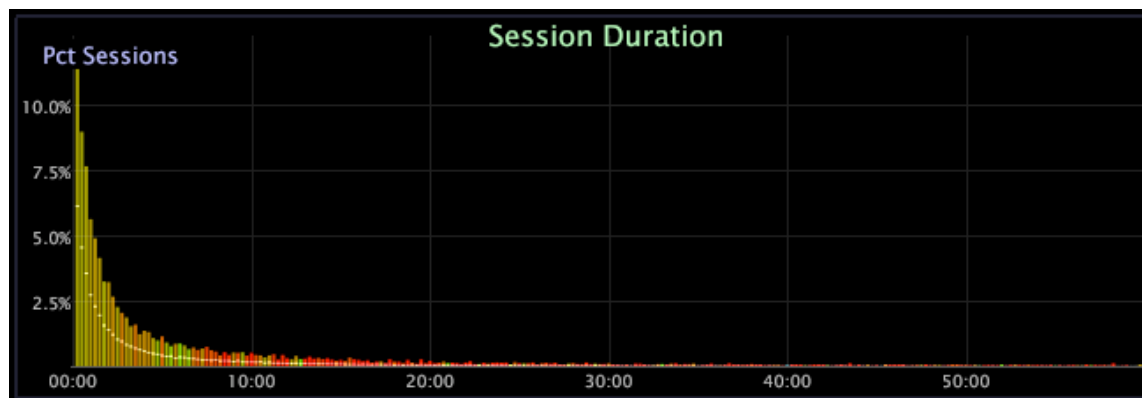


Figure 2: Typical session duration distribution for most sites, excluding 0 to 15 seconds which typically accounts for as much as half of traffic on a session-by-session basis.

Despite the metric’s general utility, there are a litany of problems associated with session duration. In the context of measuring engagement the fundamental problems with using session duration are that *you have no way to know if the visitor was actually paying attention to your site or not.*

Consider these examples:

- (1) The visitor loads your home page and within 20 seconds clicks to a product page and begins reading. Then, the phone rings and they spend the next 28 minutes talking, not paying attention to your site. *Total session duration: 28 minutes and 20 seconds.*
- (2) The visitor loads your home page as well as three other sites in multiple browsers. The visitor spends a combined hour browsing all four sites, switching back-and-forth between browsers. *Total session duration: One hour.*
- (3) The visitor comes to your site through Google after searching for a specific product and within 15 seconds finds exactly what they are looking for on the page they entered on. *Total session duration: Not reported.*

Which example best describes an “engaged” individual? Based on session duration, you would probably be tempted to say the second one—surely an hour session duration means a well-engaged visitor, right? However, it is likely that the most “engaged” visitor based on measured behavior is probably the third one. Unfortunately session duration isn’t reported for single-page sessions so duration as a measure of engagement fails.

Considered broadly, session duration is a useful indicator of the general propensity of your audience to view your site and is very useful when examining the distribution of visitors by session duration (for example, the percentage of visitors with session

durations under 90 seconds.) But as a functional measure of engagement, session duration alone is not robust enough to accurately describe the visitor's involvement, interaction, and Attention.

PAGE VIEWS PER SESSION

Page views per session is defined by the Web Analytics Association as “the number of page views in a reporting period divided by number of visits in the same reporting period” without qualification (Figure 3.) Page views per session attempts to measure engagement as a function of the total number of clicks generated and the total amount of content viewed. While this metric is perhaps less ambiguous than session duration, how would you score the engagement of the following visitors:

- (1) The visitor searches Yahoo for your company's name and “customer service department for complaints” which brings them directly to your contact page. Then they click on a link that opens a mail dialogue in Outlook. *Total page views per visit: one page.*
- (2) The visitor comes directly to your site and randomly reads 30 pages of content but never downloads or otherwise provides information and never returns to your site. *Total page views per visit: 30 pages.*

On the surface the second visitor would appear to be most engaged. The problem is that the first visitor knew exactly who you were and exactly what they wanted when they came to your site and, upon quickly finding the information they needed, directly engaged your customer service staff with a problem.

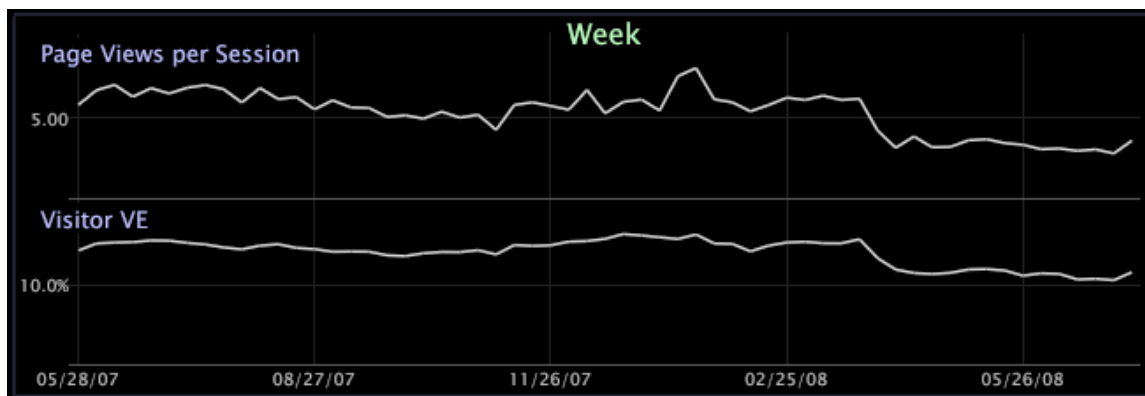


Figure 3: Typical view of page views per session compared to Visitor Engagement. Note that while page views per session declined as much as 50% between March and April, Visitor Engagement was only down roughly 20% during the same period.

Page views per session is similar to session duration, useful as a general indicator and more valuable when appropriately bucketed (for example, the percentage of sessions having more than 20 page views.) But as a measure of visitor engagement, page views per session simply does not tell enough about what the visitor was thinking or doing while viewing those pages.

VISITS PER VISITOR

Oddly not covered by the Web Analytics Association, visits per visitor are often used to understand the loyalty, frequency, and recency of a visitor to a site over time (Figure 4.) The metric is basically a count of the total number of times a visitor has visited the site, although ideally it is calculable over whatever timeframe the analysis requires.

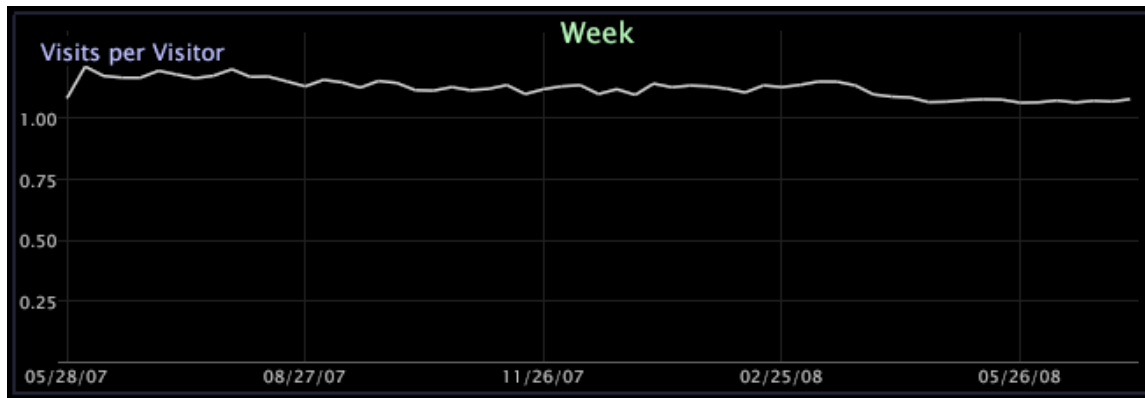


Figure 4: Typical graph of visits per visitor accounting for the fact that most visitors never pay enough Attention to become truly engaged.

When using visits per visitor to measure engagement there is a natural tendency to equate more visits with greater engagement. While that may be true, this thinking fails to capture the enthusiasm for your products or brand that a visitor new to your site may have. Consider these examples:

- (1) A visitor who has never been to your site before finds you through a Google search for your company name, views dozens of pages over 45 minutes, requests more information after downloading three documents, and subscribes to your corporate weblog. *Visits per visitor: one visit.*
- (2) A visitor who has only been to your site three times but each time posted a very positive review of your products and commented on several blog posts. *Visits per visitor: three visits.*
- (3) A visitor who has come to your site every week for the last 20 weeks but only looks at a single page and has never requested more information, downloaded a document, or subscribed to any of your RSS feeds. *Visits per visitor: 20 visits.*

While less ambiguous than the time and click-depth metrics, visits per visitor is difficult to use as a measure of visitor engagement. The metric is improved by examining the frequency of visitation (for example, percent of visitors visiting at least three times per month) or recency of visit (for example, percent of visitors visiting within the past week) and is certainly useful to determine the overall propensity to return of your visiting audience.

CONVERSION RATE

Conversion rate is simply defined by the Web Analytics Association as “a visitor completing a target action” and expanded on to describe how conversion can be used to segment visitors. Many people consider conversion to be the ultimate measure of engagement. We disagree.

While the completion of key activities on the web site is undoubtedly of fundamental importance to determining the overall engagement of an individual, this measure alone is insufficient to accurately describe visitor engagement, especially over long periods of time. Consider the following:

- (1) A visitor to a retail web site who spends hours browsing different products, reading consumer generated reviews, and comparing prices but who fails to add a single item to a shopping cart, much less make a purchase. *Conversion rate: Zero percent for purchase.*
- (2) A visitor to a marketing site who downloads a document but never returns to the site. *Conversion rate: 100 percent for downloading.*
- (3) A visitor to a blog who subscribes to the blog's RSS feed and then returns to the site 100 times in the next year. *Conversion rate: 1 percent for blog subscription.*
- (4) A visitor to a retail web site who visits the site 10 times prior to making a purchase. *Conversion rate: 10 percent for purchase.*

While we agree that conversion rate is critical to understanding visitor behavior particularly the reaction of visitors to specific messages, conflating conversion and engagement fails to recognize that ***not every session is necessarily an opportunity to convert a visitor.***

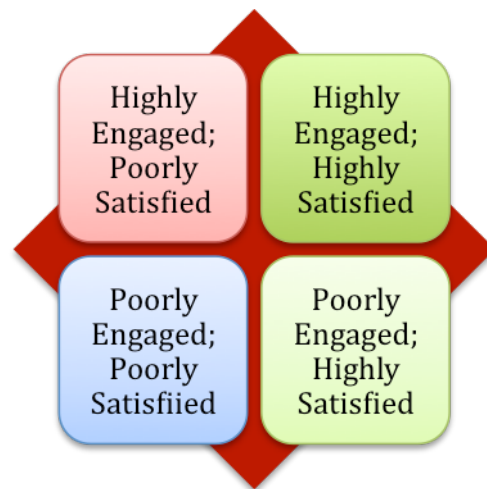
The Visitor Engagement calculation described in this document can help business analysts better understand the behavior of the 95% who are not making a purchase. Considering the widely held belief that the best customers are the ones you already have, engagement data used wisely may open the door to all kinds of new and valuable customer interactions, if only the business is ready and able to look.

CUSTOMER SATISFACTION

Of all the metrics presented, customer satisfaction is perhaps the closest to engagement in terms of its calculation and use. Typically determined by asking visitors a series of qualitative questions about their experience with the site, products, and brand - then algorithmically translating those responses into a quantitative score -- customer satisfaction calculated rigorously is able to help identify nuanced differences in the visitor experience.

But satisfaction is not engagement; satisfaction is complimentary to engagement. Consider the following:

- (1) A visitor comes to your site for the first time ever, browses a few pages, and is then asked to complete a satisfaction survey in which they rate the site and brand favorably. The visitor never returns to the web site. *Customer satisfaction: Likely very high.*
- (2) A visitor comes to your site angry about a product they recently purchased and spends 30 minutes trying to find a solution to their problem prior to going to the contact form and sending you an angry message and completing a satisfaction survey. *Customer satisfaction: Likely very low.*
- (3) A visitor comes to your site, browses dozens of pages over 40 minutes and completes several important activities as defined by your business stakeholders. When asked to complete a satisfaction survey, she declines. *Customer satisfaction: Unknown.*



Customer satisfaction and visitor engagement are complimentary, each designed to measure something specific about the ongoing relationship the visitor has with the site, company, or brand. Both measures are a necessary component of a site's ongoing effort to better understand visitors as well as refine technology and marketing efforts to best serve *both* the audience *and* the business.

Larry Freed, CEO of one of the leading customer satisfaction measurement firms and a respected blogger, commented in October 2007 ⁽⁷⁾ that “the visitor engagement calculation will not measure success” and that success is a measure of whether the users accomplished what they wanted and whether they were satisfied. We agree with Mr. Freed—Visitor Engagement is not designed to measure success; conversion rate is designed to measure success. And visitor engagement is not designed to measure whether visitors were satisfied; the customer satisfaction metric does that perfectly well. ***The Visitor Engagement metric is designed to allow site operators to measure the multitude of interactions that are neither successful nor satisfying but comprise the largest part of the measurable interactions occurring online.***

It is important to note that the authors do not fault any reader for using the above metrics as a proxy measure of engagement. Understanding and measuring Visitor Engagement is both relatively new and can be moderately complex. Some site owners haven't given the question of “what defines an engaged visitor” nearly enough thought; others haven't yet deployed powerful enough technology to make the necessary calculation. But for those readers who have given this idea significant thought and are still looking for answers, we propose the following framework for measuring engagement.

⁷ http://www.freedyourmind.com/freed_your_mind/2007/10/on-engagement.html

THE VISITOR ENGAGEMENT CALCULATION

The Web Analytics Demystified Visitor Engagement calculation is designed to allow the scoring of individual visitor behavior against pre-determined criteria. The calculation accommodates the major proxies for success described in the previous section—session duration, page views per visit, visits per visitor—and yields a single measure that can be applied to any measured dimension in your web analytics reporting toolkit. More importantly, the resulting metric can be used in compliment to other measures of success online including conversion and customer satisfaction to improve the businesses overall understanding of visitor and customer behavior.

The Visitor Engagement calculation is described as follows:

$$\Sigma(C_i + D_i + R_i + L_i + B_i + F_i + I_i)$$

While the calculation may look complicated, each of the component indices are actually very straight forward and leverage the most basic web analytics data and calculations. What results is a percentage score of visitor engagement composed of a series of individual indices (Figure 5.)

Visitor Email	Ci	Di	Ri	Li	Ii	Bi	Fi	Visitor VE
judahphillips	100.0%	40.0%	83.3%	100.0%	0.0%	80.0%	0.0%	43.6%
stefan.backl	0.0%	0.0%	100.0%	0.0%	0.0%	20.0%	0.0%	16.8%
veronique@v	0.0%	0.0%	100.0%	0.0%	0.0%	25.0%	0.0%	17.9%
eric.peterson	100.0%	0.0%	66.7%	0.0%	100.0%	100.0%	0.0%	40.8%
g.boot@ned	66.7%	0.0%	21.4%	0.0%	0.0%	66.7%	0.0%	21.1%
amity.peters	66.7%	0.0%	75.0%	0.0%	33.3%	66.7%	0.0%	40.2%
r.deloumeau	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	4.3%
jim@jimnovo	0.0%	0.0%	37.5%	100.0%	0.0%	0.0%	0.0%	20.1%
brosnaro@st	0.0%	0.0%	37.5%	0.0%	0.0%	33.3%	0.0%	10.7%

Figure 5: Example Visitor Engagement set-up showing component indices and Visitor Engagement score (Visitor VE) for individual visitors to a web site.

NO WAY! THERE IS NOT ONE UNIVERSAL DEFINITION OF ENGAGEMENT!

Yes, yes there is. While we appreciate the argument that different businesses will all need a unique measure of engagement, we disagree. While this position requires some backtracking for Mr. Peterson, Mr. Carrabis has been eloquent and clear on this point all along ⁽⁸⁾ the idea that every company will need a different engagement calculation is like saying that every car will need a different measure of velocity or fuel consumption.

⁸ <http://www.allbusiness.com/marketing-advertising/marketing-advertising/10172624-1.html>

The analogy is apt. The Visitor Engagement calculation works much the same as measures of fuel efficiency in automobiles. While there are slight differences under the hood between vehicles, the resulting measure means the same thing regardless. The “slight differences” in the engagement calculations involve the setting of thresholds in each of the indices, far less tweaking than likely goes into the engines of full-sized pickup trucks in an effort to eek out every last mile per gallon as gas prices climb.

The tweaking of thresholds is what makes the calculation useful and relevant to different types of businesses online. Search engines have different thresholds than content sites like CNN, MSNBC, and The Wall Street Journal, just as social networks may or may not have different thresholds than online mail sites like Gmail, Yahoo Mail, and Hotmail. The thresholds described below *have to be set thoughtfully* but we do not believe that the setting of these thresholds somehow negates the broad applicability of the calculation we have described.

Furthermore, using the more advanced form of this equation (presented in the Appendix), we believe that our measure of engagement is universal and can be applied to any device, any operating system, and technology capable of outputting data regarding the level of Attention being paid.

THE FUNDAMENTAL ASSUMPTIONS

Observant readers will note that many of the following component indices make fundamental assumptions, especially when it comes to setting thresholds. Remember that the relevant definitions of visitor engagement used in this document are as follows:

Engagement is an estimate of the depth of interaction against a clearly defined set of goals.

Mr. Carrabis’s variation:

Engagement is the demonstration of Attention via psychomotor activity that serves to focus an individual’s Attention.

Attention is a behavior that demonstrates that specific neural activity is taking place.

Given these definitions, the underlying assumption is that as long as the visitor is maintaining a measurable level of interaction with the web site (paying *Attention*) we will consider them “engaged.” This measure of engagement is taken against a set of pre-determined goals to better allow users to differentiate levels of engagement among visitors.⁹ A second assumption is that each component index has an “ideal value” that when exceeded indicates a higher level of Attention.

To be clear: these are fundamentally important assumptions. Fortunately, should it be shown that Attention is not well correlated to neural activity, or that traditional web analytic measures like session duration or click-depth are poor measures of Attention, this calculation allows metric variability – a multiplicity of measurements, measuring systems and tools – so that a satisfactory metric can be derived. More importantly, the following component metrics *are* good measures of Attention, worthy of inclusion in the Visitor Engagement framework.

⁹ Note that the metric defined here uses discrete values for computational purposes. The modification shown in the Appendix allows for continuously variable values as would be used in a rolling, real-time determination of engagement.

THE TOOLS YOU NEED TO MAKE THIS CALCULATION

In general, to derive and use this metric, your web analytics tool or data warehousing environment need to have the following functionality:

REQUIRED

- The ability to create custom metrics using mathematical operators
- The ability to add new metrics to existing dimensions (for example, pages, content groups, campaigns, and referring domains)
- A disaggregated view of the visitor over time

DESIRED

- The ability to compare the elements in a dimension to a pre-determined string (for example, the *search phrase* dimension filtered against your company's brand names)
- The ability to track non-HTML content consumption, or at least the intention of the visitor to use non-HTML content (for example, RSS feeds)

While not appearing complex, unfortunately some applications like the very popular Google Analytics do not have most of these capabilities. But many of the web analytics vendors provide advanced analysis tools capable of making the calculation including Omniture Discover OnPremise, IndexTools Rubix, WebTrends Score, Coremetrics Explore, and Unica Affinium NetInsight. Furthermore, some companies are experimenting with taking a visitor-level extract from their web analytics solution, using business intelligence and data warehousing tools to make the calculation. Alternatively, readers may want to contact Mr. Carrabis's company, NextStage Global, who has provided Engagement and Attention metrics to clients since 2001.

THE COMPONENT METRICS EXPLAINED

The Visitor Engagement calculation is as follows:

$$\Sigma(C_i + D_i + R_i + L_i + B_i + F_i + I_i)$$

Since the original publication of this work, Mr. Carrabis has revised this calculation to make it mathematically correct as well as to extend the calculation's ability to be both *extendable* and *extensible* for uses not previously imagined by Mr. Peterson. An explanation and the result of Mr. Carrabis's work is found in the Appendix. The net result is that the equation above has been shown to work perfectly well when applied to a single source of information (for example, web server log files, mobile phone transaction logs, etc.)

As previously mentioned, the component indices used in this calculation are:

- **Click Depth Index:** Captures the contribution of page and event views
- **Duration Index:** Captures the contribution of time spent on site
- **Recency Index:** Captures the visitors "visit velocity"—the rate at which visitors return to the web site over time
- **Brand Index:** Captures the apparent awareness of the visitor of the brand, site, or product(s)
- **Feedback Index:** Captures qualitative information including propensity to solicit additional information or supply direct feedback
- **Interaction Index:** Captures visitor interaction with content or functionality designed to increase level of Attention the visitor is paying to the brand, site, or product(s)
- **Loyalty Index:** Captures the level of long-term interaction the visitor has with the brand, site, or product(s)

In the following sections, each of these component indices is detailed and discussed in terms of how different types of sites might set the necessary thresholds.

THE CLICK DEPTH INDEX

The Click-Depth Index (C_i) proxies Attention as a function of click-depth on a session-by-session basis. Based on the commonly used page views per session metric, the Click-Depth Index resolves noise caused by visitors bouncing off the site after viewing only a small number of pages by only scoring the session when click-depth exceeds a pre-determined threshold.

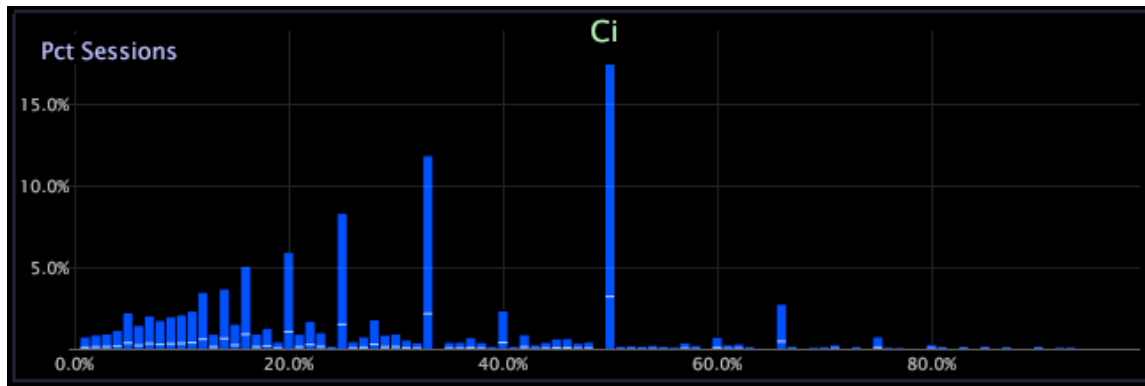


Figure 6: Distribution of Ci values for over 5,000 visitors over 27,000 sessions.

For example, if analysis revealed that the largest percentage of visitors who were making purchases were viewing at least 4 pages during their sessions, a reasonable threshold for the Click-Depth Index would be “4 pages viewed”.

The resulting calculation would be:

$$C_i = \text{Sessions having "at least 4 page views"} / \text{All Sessions}$$

The simplest example of this calculation would be for a single visitor over four separate sessions:

1. Session #1: Visitor has 3 page views: $C_i = 0$
2. Session #2: Visitor has 8 page views: $C_i = 1$
3. Session #3: Visitor has 1 page view: $C_i = 0$
4. Session #4: Visitor has 12 page views: $C_i = 1$

The total Click-Depth Index for this visitor would then be:

$$C_i = (0 + 1 + 0 + 1) / 4 \text{ total Sessions} = 50\%$$

When this calculation is applied to relevant dimensions in your reporting package, you will see results similar to the data presented in Figure 7.

Referring Domain	Ci	Sessions
webanalyticsassociation.org	40%	238
blogspot.com	27%	209
yahoo.com	26%	720
google.com	25%	5,441
wikipedia.org	15%	196

live.com	14%	345
google.co.uk	14%	778
msn.com	6%	172
images.google.fr	4%	163
stumbleupon.com	0%	957

Figure 7: Sample data for the Click-Depth Index reported against the referring domain dimension.

In Figure 7, while Google originated 5,441 sessions to this site, only 25% of all of those sessions exceed the click-depth threshold and were scored as such. Compare the Click-Depth Index to the commonly reported metric average page views per session:

Referring Domain	Ci	Page Views per Session
webanalyticsassociation.org	39.90%	9.03
blogspot.com	26.80%	5.89
yahoo.com	26.00%	5.25
google.com	24.80%	5.12
wikipedia.org	15.30%	4.18
live.com	14.20%	4.45
google.co.uk	13.90%	3.85
msn.com	5.80%	2.76
images.google.fr	3.70%	2.56
stumbleupon.com	0.40%	2.08

Figure 8: Comparison of the Click-Depth Index (Ci) and page views per session, reported against referring domain.

Here the two metrics trend in the same direction, while average page views per session describes the entire audience, the Click-Depth Index tells you something more specific. In Figure 7, you can see that the Web Analytics Association sent 238 sessions to the site and the “average” session was 9 page views in depth (Figure 8). Interesting, but less interesting than knowing that 40 percent of those sessions were at least 5 page views deep. The former metric provides a view of your entire audience; the latter tells you about a specific segment of your audience (those visitors having sessions deemed to be engaged from a click-depth perspective.)

Conversely, MSN.com, the French Google Images domains, and Stumbleupon are all sending traffic having approximately 2 page views per session so this metric offers very little resolution. However, the Click-Depth index shows a much higher level of differences between these three sites, highlighting that there is likely very little attention being paid by visitors from Stumbleupon ($C_i = 0.40\%$) but slightly more by MSN.com visitors ($C_i = 5.80\%$)

THE DURATION INDEX

The Duration Index is similar to the Click-Depth Index: Attention as a function of session duration.

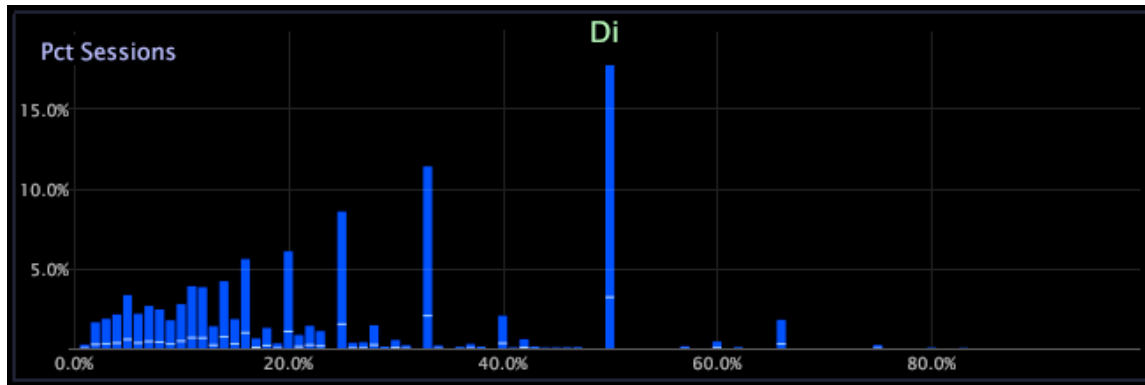


Figure 9: Distribution of Di values for over 5,000 visitors over 27,000 sessions.

If you believe that visitors are paying Attention to your site when they spend more than five minutes reading content, you would set the threshold for the Duration Index to “5 minutes” and the resulting calculation would be:

$$D_i = \text{Sessions where "duration over 5 minutes"} / \text{All Sessions}$$

Some have commented that the Click-Depth Index and Duration Index are measuring the same thing, making one extraneous. But consider the following examples:

- A visitor comes to your site and in less than five minutes navigates through twelve pages to the information they are looking for. *Session duration: Under 5 minutes; click-depth: 12 pages.*
- A visitor comes to the site and spends 20 minutes reading content on three pages. *Session duration: 20 minutes; click-depth: 3 pages.*
- A visitor comes to the site and spends only one minute reading three pages. *Session duration: 1 minute; click-depth: 3 pages.*
- A visitor comes to the site and spends 40 minutes reading dozens and dozens of pages. *Session duration: 40 minutes; click-depth: dozens of pages.*

The first and second visitors appear to be paying the most Attention based on their content consumption but each in different ways; the third visitor seems only poorly engaged (at least using these two indices) and the fourth appears to be well-engaged. If you were to pick only either the Click-Depth or Duration indices, either the first or second visitor would be under-represented in the calculation. Even in small samples you begin to see where these two indices converge (Figure 10).

Referring Domain	Di	Ci
webanalyticsassociation.org	20%	40%
blogspot.com	20%	27%
google.com	17%	25%
yahoo.com	16%	26%
wikipedia.org	14%	15%
google.co.uk	14%	14%
live.com	9%	14%
msn.com	8%	6%
stumbleupon.com	7%	0%
images.google.fr	7%	4%

Figure 10: Sample data for Duration Index (Di) shown with Click-Depth Index (Ci) against referring domains.

In the example, while only 20 percent of visitors from the Web Analytics Association site are exceeding our duration threshold, 40 percent are exceeding our click-depth threshold so are more like the first visitor described in the example above. Visitors from the Wikipedia are equally likely to exceed both thresholds; visitors from Stumbleupon are far more likely to exceed the duration threshold despite failing to click very deeply into the site.

THE REGENCY INDEX

The Recency Index measures the likelihood that the visitor has been to the site multiple times in the recent past. Its inclusion in this framework is credited to the great work Jim Novo has done studying how recency of visit affects customer behavior⁽¹⁰⁾. The logic behind recency is very simple—according to Novo, recency is “the number one most powerful predictor of future behavior” and “the more recently a customer has done something, the more likely they are to do it again.”

¹⁰ <http://www.jimnovo.com/Recency-Model.htm>

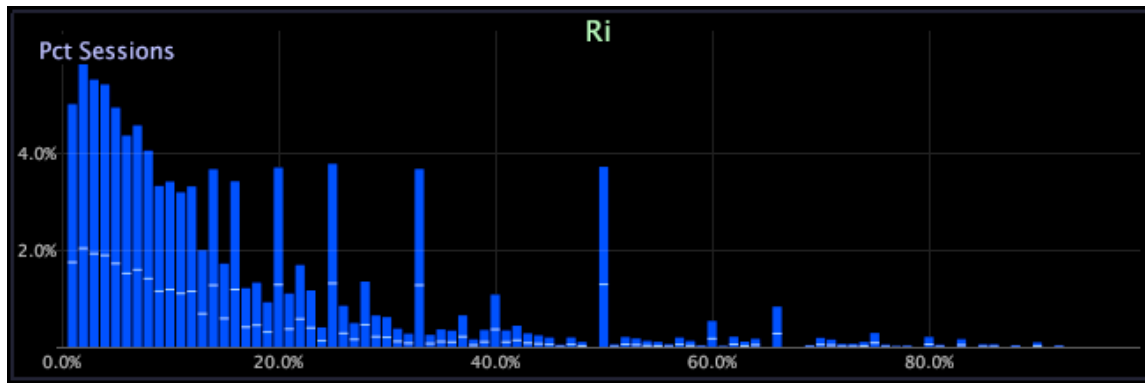


Figure 11: Distribution of Ri values for over 5,000 visitors over 27,000 sessions.

The Recency Index examines the visitor's "visit velocity"—the rate at which they return to the site. The calculation is very simple for any given session:

$$Ri = 1/\text{Number of days elapsed since the most recent session}$$

For example, if the visitor came to the site on 16 May 2008 (Visit #1) and again on 18 May 2008 (Visit #2), then their Ri score would be as follows:

$$Ri = \{1/(18-16)\} = 1/2 = 0.50$$

In this case the average Ri score for the visitor would be 0.25 (25%):

$$Ri_{\text{cum}} = Ri / 2 \text{ Sessions} = 0.50 / 2 = 0.25 (25\%)$$

Recency does not require any threshold and simply scores against the time elapsed since the last visit. Note that visitors who come to the site frequently will have high Recency Indices owing to relatively small denominators for many of their visits. Conversely, infrequent visitors will have low Recency Indices for the opposite reason.

THE LOYALTY INDEX

The Loyalty Index is an Attention proxy in that it recognizes that any visitor who has come to your site repeatedly over time is very likely paying Attention, at least to a small degree. The basic calculation for the Loyalty Index is as follows:

$$Li = 1 - (1 / \text{Number of visitor sessions during the timeframe})$$

The resulting measure ranges between 0% for those visitors who have only been to the web site a single time to nearly 100% for those visitors having a very high number of recorded sessions during the timeframe under examination (Figure 12).

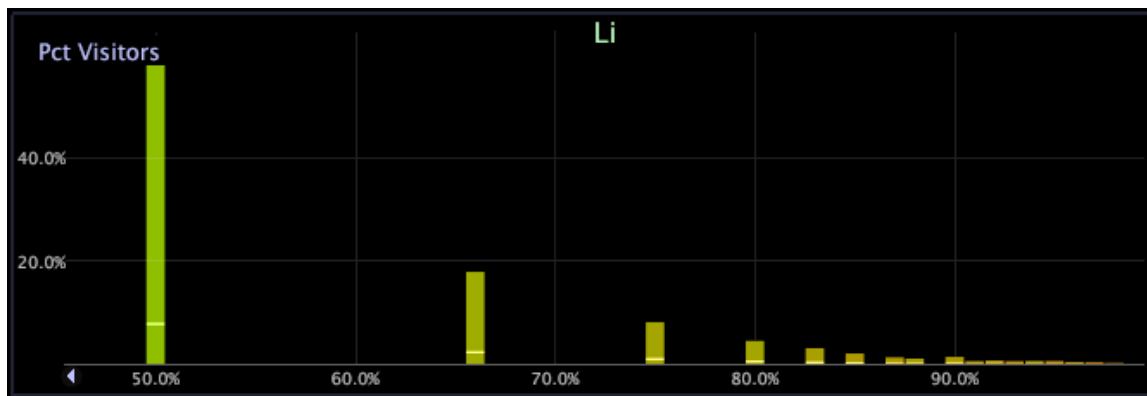


Figure 12: Distribution of Li values for over 5,000 visitors over 27,000 sessions. The large groups occur because of the significant number of visitors in the data set that have only visited a small number of times.

What is important to recognize is that this variable can be either binary (someone is loyal or they're not) and, as shown in the Appendix, can take on a scale series that allows several layers of "loyalty" to be determined. This allows companies to determine what events are required for greater loyalty and steer visitors to those events.

It is also worth noting that "loyalty" used in this context is different than other definitions of loyalty. In more common usage, loyalty is often measure of the fidelity and faithfulness a person has to another person, object, or cause. For example, a searcher "loyal" to Google would *never* use Yahoo, Ask, or MSN to search the Internet in the same way a "loyal" spouse would never cheat. Marketers often refer to loyalty in terms of share of requirements—the percentage of brand purchases relative to the total number of category purchases by brand buyers⁽¹¹⁾.

¹¹ <http://www.whartonsp.com/articles/article.asp?p=463943&seqNum=6>

THE BRAND INDEX

The Brand Index proxies the level of Attention that the visitor is paying to your brand just prior to reaching your site. Measured by examining all of the inbound search phrases to your site from search engines like Google, Yahoo, and MSN, then comparing those phrases to branded keywords, the Brand Index helps you infer whether or not the visitor was already thinking about *you and your products or services specifically* as they began their involvement and interaction with your site.

Clickthru Search Phrase	Bi	Visitor VE
web analytics demystified	100.0%	18.5%
kpi	2.5%	5.0%
eric peterson	100.0%	17.7%
web analytics wednesday	7.0%	13.2%
key performance indicators	1.3%	2.6%
key performance indicator	1.6%	4.0%
web analytics	3.3%	8.4%
web analytics jobs	6.4%	14.0%
webanalyticsdemystified	100.0%	19.1%
web analytics forum	9.6%	14.9%
web analytics wednesdays	8.6%	13.6%
site:www.webanalyticsdemystified.com index	100.0%	14.3%
judah phillips	13.6%	15.7%

Figure 13: Sample of search phrases used to find Web Analytics Demystified and their associated Bi and Visitor Engagement scores. Branded terms will always have a 100% value for Bi; non-branded terms Bi values are a function of the Visitor Engagement for visitors who have used those terms.

Calculating the Brand Index requires that your analytics application provides a programmatic way to evaluate and score each session originating from search based on the search phrase the visitor used. To make the calculation, you first need to make a reasonable list of branded terms associated with your company. As an example, Apple Computers might come up with the following list:

Apple, Apple Computers, iPod, iPhone, iMac, iTunes, iLife, iWork, Steve Jobs, Mac Pro, OS X, nano, touch, Leopard

You may also want to consider incorporating direct traffic into your Brand Index, meaning that you would score the session *either* when the search phrase matched your branded list *or* when there was no referrer for the session. *The Brand Index does not require any threshold other than the creation of a set of branded terms used for scoring the session.*

The logic behind the Brand Index is simple. If you have created an impression on someone such that they are searching for your products, services, or company, or if you have already demonstrated value such that people either know your site's URL or have bookmarked it, *then that visitor is already paying Attention!* That Attention may not have translated into a high level of engagement yet, but it is difficult to argue that someone searching directly for you is just casually drifting around the Internet. Further, an individual performing the activities described here has de facto demonstrated engagement in the past (to create such an impression the individual must have been engaged as a function of engagement is to focus awareness on and store information in memory).

It is very likely that visitors coming to a site via a "branded" search already have some belief in place about a product or service and the object of their visit is to confirm or deny the pre-existing beliefs. The fact that these beliefs exist and that the object of

their visit is to confirm or deny those beliefs indicates that a pre-existing level of “engagement” exists between the visitor and the brand. This pre-existing level of “engagement” directly affects all other variables being considered in this calculation.

For example, the searcher’s navigation will be neither random nor a “blind search” pattern (one in which information is desired but the searcher doesn’t know exactly what information they desire, so search “blindly”): it will be highly targeted towards confirming or denying their pre-existing beliefs. The initial formulation of the Brand Index shown in the Appendix and the expanded final form of the calculation demonstrate this methodology and allow the final calculation to be made for any number of “branded” and “unbranded” sessions, even allowing for various degrees of branding to be determined.

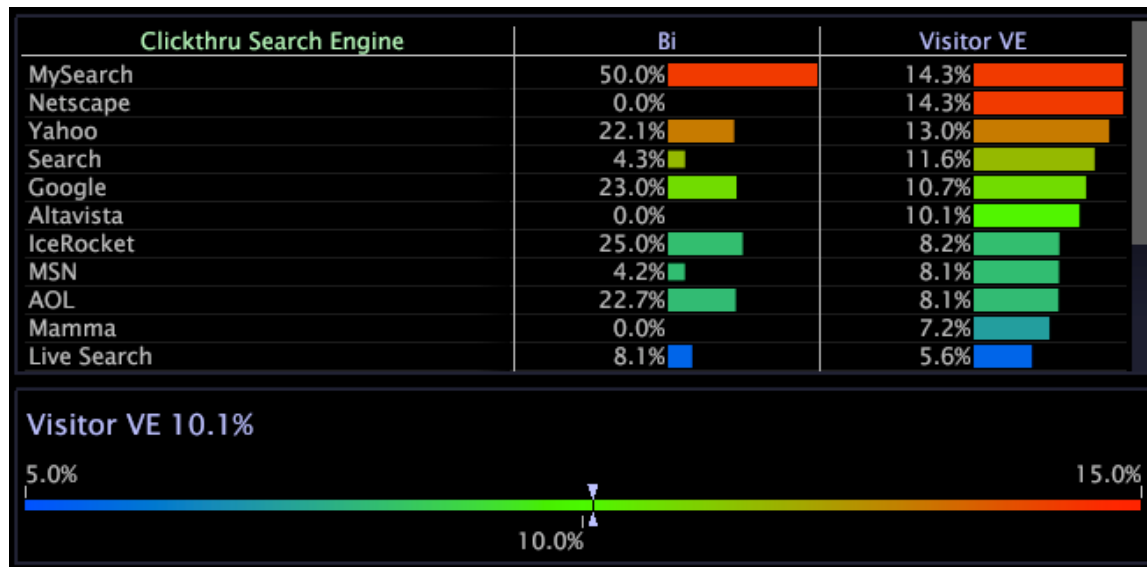


Figure 14: Sample of search engines driving traffic to Web Analytics Demystified showing their Bi and Visitor Engagement scores.

THE FEEDBACK INDEX

The Feedback Index is the primary qualitative input to the Visitor Engagement calculation. As such, this index often requires a more robust data collection strategy than is provided by most web analytics packages, essentially the integration of qualitative data collected using feedback, CRM, or customer satisfaction applications. Fortunately, there is a way for all sites to use this index with little or no modification to their data collection strategy.

The Feedback Index has a tendency to report very low scores on a site-wide basis but does provide valuable insight into the level of engagement on a visitor-by-visitor basis (Figure 15.) *No threshold is required for the Feedback Index*—every session in which feedback is gathered is scored positively.

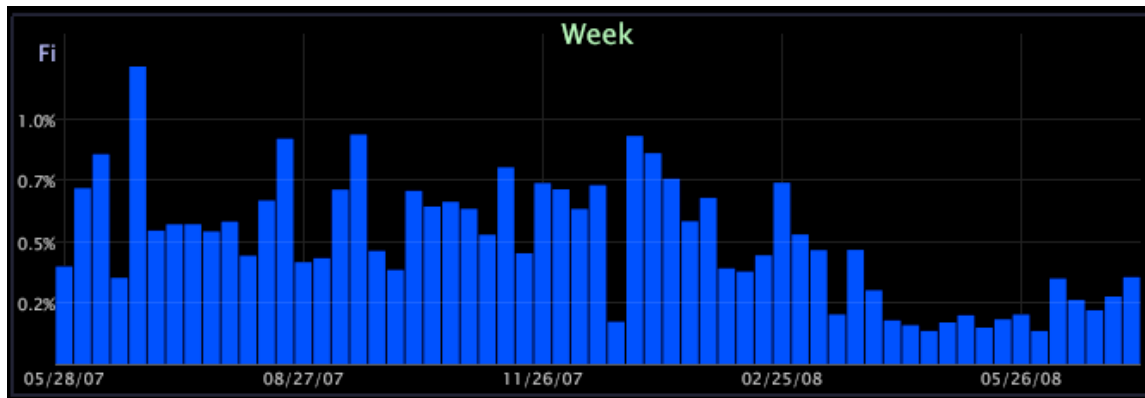


Figure 15: Fi scores trended on a week-over-week basis for over 5,000 visitors over 27,000 sessions.

FEEDBACK INDEX: THE EASY WAY

The easy way to calculate the Feedback Index is to simply score any session in which a visitor provides any type of direct feedback through a measurable mechanism. Common examples of such mechanisms found on the Internet include:

- Email links
- Feedback forms
- Click-to-call links
- Page feedback and rating tools

Assuming that any of these mechanisms found on your site are properly tagged, the Feedback Index calculation becomes:

$$F_i = \text{Sessions where visitor submits feedback} / \text{All Sessions}$$

FEEDBACK INDEX: THE LESS EASY WAY

The “less easy” strategy to calculate the Feedback Index is to leverage third-party applications designed to capture qualitative customer feedback or power online surveys and use that data as the basis of the feedback calculation. While somewhat more involved from a collection and data integration standpoint, this type of “Voice of Customer” data is fundamental to every site’s ability to truly know their visitors and is a critical component of the larger Web Site Optimization Ecosystem⁽¹²⁾.

Assuming you’re able to gather Voice of Customer data, the Feedback Index calculation can become much more robust in terms of measuring intimacy and influence. With the ability to ask customized questions, you can then make your calculation based on the responses to questions like:

- “Do you have a generally positive opinion of our brand?”
- “Would you describe yourself as passionate about our services?”
- “Would you recommend our products or services to a friend or family member?”
- “Have you written a positive review of one of our products recently? What about a negative one?”
- “Are you likely to use our service again in the future?”

¹² http://www.foreseerresults.com/Form_Epeterson_WebAnalytics.html

These purely qualitative questions all lend themselves to quantitative responses that can be scored as part of the Feedback Index. For example:

$$F_i = \text{Number of positive responses} / \text{All Qualitative Questions Asked}$$

The simplest example of this type of Feedback Index would be one in which you asked the visitor four questions regarding their awareness of your site/brand/products, two of which they respond positively and two either negative or neutral, yielding a Feedback Index of 50 percent.

$$F_i = 2 \text{ positive responses} / 4 \text{ total responses} = 50\%$$

If you choose this strategy, both the “easy” and “less easy” approaches should be included your calculation, making your Feedback Index something like:

$$F_i = ((\text{Sessions where visitor submits feedback} / \text{All Sessions}) + (\text{Number of positive responses} / \text{All Qualitative Questions Asked})) / 2$$

This strategy does involve the integration of your Web Analytics and Voice of Customer platform. Consult your web analytics vendor for specifics about how to enable this integration and make the resulting Feedback Index calculation. Additionally, Mr. Carrabis has expanded on the use and calculation of the Feedback Index in the Appendix.

THE INTERACTION INDEX

The Interaction Index measures exactly that: specific actions that visitors take while engaging with your site. This index is the most flexible described by this framework and the one requiring most thought and customization on the part organization’s using the Visitor Engagement metric.

Engagement Goals	Pct Highly Engaged	Visitor VE
Downloaded a PDF	28.8%	21.2%
Researched a Web Analytics Job	24.9%	21.2%
Engaged a Web Analytics Wednesday Host	29.0%	21.8%
Researched Posting a Web Analytics Job	35.8%	23.7%
Participated in Web Analytics Wednesday	17.3%	20.5%
Hosted a Web Analytics Wednesday	43.9%	23.8%
Submitted a Comment to a Blog	38.9%	24.8%
Shared a Social Bookmark	27.5%	20.8%

Figure 16: Example engagement goals measured via the Interaction Index. Visitors participating in these events are scored as “paying Attention” and thus their Visitors Engagement scores are relatively high (the average VE score for this site is around 10%)

Similar to the Brand Index, the Interaction Index gives you a programmatic way to score desired actions on the part of visitors that you believe to indicate engagement. This is not to say that you need to conduct a full statistical regression analysis to determine which actions are positively correlated with engagement—web analytics is simply not that precise. Rather it is a way

to capture those actions and events that your organization believes are strong indicators of engagement, either with the brand or with the web site.

No threshold is required for the Interaction Index. Like the Brand Index, the Interaction Index will score positively if the visitor completes any of the defined interactions during the session.

Examples include:

- Submitting a user review of a product
- Submitting a blog comment
- Saving a page or post in a social bookmarking tool like Digg or Flickr
- Downloading a PDF
- Viewing a video or listening to a podcast
- Using a Rich Internet Application (RIA)
- Adding an item to a shopping cart

The calculation for the Interaction Index depends heavily on how your data is being collected and reported and the capabilities of your reporting package. The essence of making this calculation is the same as the Brand Index—you form a list of events and/or URLs that identify the action being measured and then score each session in which at least one of those events occurs:

$$I_i = \text{Sessions where visitor completes an action} / \text{All Sessions}$$

The best strategy for coming up with a *reasonable number of actions* to track using the Interaction Index is simple: ask people. Ask people in your organization what they think visitors and customers would be doing on your web site if they were actively paying Attention to you. More often than not, you'll get a perfectly reasonable list of answers, all of which are more-or-less easily scored. Even in very large companies you should be able to get a list of activities deemed most engaging on a business unit by business unit basis, and *it is okay for each business unit to have their own visitor engagement calculation.*

If you get an unwieldy list you may want to consider spending some time re-evaluating your business objectives. No web site can be all things to all people, and any attempt to make this happen is likely an indicator of a poorly designed web site.

SETTING THRESHOLDS FOR CLICK-DEPTH AND DURATION INDICES

One issue that readers are likely to struggle with is the need to set thresholds for the Click-Depth and Duration Indices. While not entirely arbitrary, the setting of these thresholds is a somewhat imprecise given that if we knew what depth of session or session duration indicated an “engaged” visitor we would have little need for the calculation presented in this document. As described above, we recommend setting the thresholds for these indices against known averages for visitors who are otherwise accomplishing a business-critical task on your site. For example:

- For a retail site, consider setting the thresholds based on the averages for all visitors who have completed a transaction in the last 90 days
- For a content site, consider setting the thresholds based on the averages for visitors who have subscribed to an email or have otherwise shared contact information or feedback with you

- For a B2B site, consider setting the thresholds based on the averages for visitors who have submitted a lead, or better, who have submitted a qualified lead
- For a social network, consider setting the thresholds based on the averages for visitors who have a substantial number of “friends” or contacts

The goal is not to be perfect with how these thresholds are being set since perfection is likely unattainable anyway; the goal is to create a reasonably high threshold against which the average visitor can be judged to determine if they are paying Attention to you. There are obviously more complex strategies for setting these thresholds—eye tracking studies, mouse-movement tracking, the use of panel-based services or other types of user activity recorders, etc.—but the goal should not be to make it more complex.

Finally, because Click-Depth and Duration are but two of seven component indices, any imprecision in setting their thresholds is somewhat mitigated by the other inputs to the calculation. Given that all of the other calculations are defined free of this type of subjective threshold, their weighting in the overall calculation will support the discovery of engaged and un-engaged individuals coming to your site.

EXCLUDING INDIVIDUAL INDICES

Not being able to make one of the index calculations is not a problem. By excluding one or more component indices, you’re just applying a “zero weighting” to the index and not changing the calculation at all.

The ideal situation is one where all component indices can be used in order to increase the overall granularity of the calculation. That said, consistency is the key: making the calculation in the absence of the Brand Index is fine as long as (if and when) you’re able to add the Brand Index, you communicate that clearly to everyone using the engagement calculation in their reporting.

PUTTING IT ALL TOGETHER: MAKING THE CALCULATION

Once you’ve successfully set up your application to calculate the individual indices and set their thresholds appropriately, the final step in making the calculation involves doing some very simple math. Recall the equation:

$$\Sigma(C_i + D_i + R_i + L_i + B_i + F_i + I_i)$$

You need to make the calculation against *whatever segment or group of visitors you’re looking at in your current report or view*. For example, if you were evaluating your referring domain dimension, you would make the calculation against *all of the visitors from each referring domain for whatever timeframe you have selected*.

Sounds difficult, but keep in mind that your web analytics application will be doing the work for you. The result will look something like Figure 17:

Referring Domain	VE	Ci	Ri	Di	Li	Bi	Fi	Ii
------------------	----	----	----	----	----	----	----	----

google.com	12%	25%	31%	17%	4%	0%	1%	8%
stumbleupon.com	9%	0%	63%	7%	0%	0%	0%	0%
google.co.uk	9%	14%	35%	14%	2%	0%	1%	3%
yahoo.com	13%	26%	31%	16%	8%	0%	1%	13%
live.com	8%	14%	33%	10%	1%	1%	1%	6%
webanalyticsassociation.org	13%	41%	29%	20%	4%	0%	0%	6%
blogspot.com	13%	27%	35%	20%	11%	1%	0%	10%
wikipedia.org	8%	15%	25%	14%	1%	0%	0%	4%
msn.com	8%	6%	50%	8%	1%	0%	0%	1%
google.ca	14%	30%	36%	20%	4%	2%	1%	9%

Figure 17: Sample of Visitor Engagement and component metrics shown against the Referring Domain dimension.

For these nine domains the average visitor engagement is just over 10 percent; those values highlighted in red indicate domains sending visitors below the average.

Examining a single domain:

Referring Domain	VE	Ci	Ri	Di	Li	Bi	Fi	li
google.com	12%	25%	31%	17%	4%	0%	1%	8%

We can see that Google performs slightly better than average in terms of bringing engaged visitors, likely driven by Interaction and Subscription indices higher than the group average. Looking further down the list, one domain appears to be nearly double the average visitor engagement for referring domains:

Referring Domain	VE	Ci	Ri	Di	Li	Bi	Fi	li
clickz.com	19%	27%	50%	33%	7%	0%	0%	27%

Here we can see that ClickZ, the popular media site, has an average visitor engagement of 19 percent, owing apparently in large part to an influx of recent visitors interacting with the site and spending a fair amount of time doing so.

Knowing that ClickZ sends visitors that are 40 percent more engaged than those from Google is interesting to the marketer and analyst, especially since Google sends a dramatically higher traffic volume. There appears to be an opportunity between the site and ClickZ, one worth exploring. Looking at “traditional” metrics for ClickZ fails to yield the same level of insight—none of the visitors converted, the average visitor had been to the site 1.7 times, viewing an average of 4.5 pages over just-under nine minutes.

Notice that we're not dealing at all with traffic volumes in the analysis above. If you want to value your traffic by volume, you already have an excellent set of metrics—page views, sessions, and visitors. We're not looking at conversion rate—although it is interesting to note that the highest converting site is one with a just-better-than-average visitor engagement value—nor are we looking at customer satisfaction.

The visitor engagement metric is designed to add value to the existing set of metrics, not act as a replacement. It is designed to quickly call our attention to domains, pages, campaigns, geographies, and other factors that are distinguished as somehow attracting larger numbers of visitors who satisfy a pre-determined set of criteria describing the level of Attention they're paying to you, your site, your products, and your brands.

EXAMPLES OF VISITOR ENGAGEMENT IN ACTION

For marketers, the operational measure of visitor engagement provides another valuable input into the process of deciding which message, which creative, which placement, which segment, and which site will provide the optimal return on investment. Especially given that not all sites have a commerce component and cannot rely on traditional measures like conversion and average order value as benchmarks for success, the Visitor Engagement metric provides a robust measure of campaign success (Figure 18).

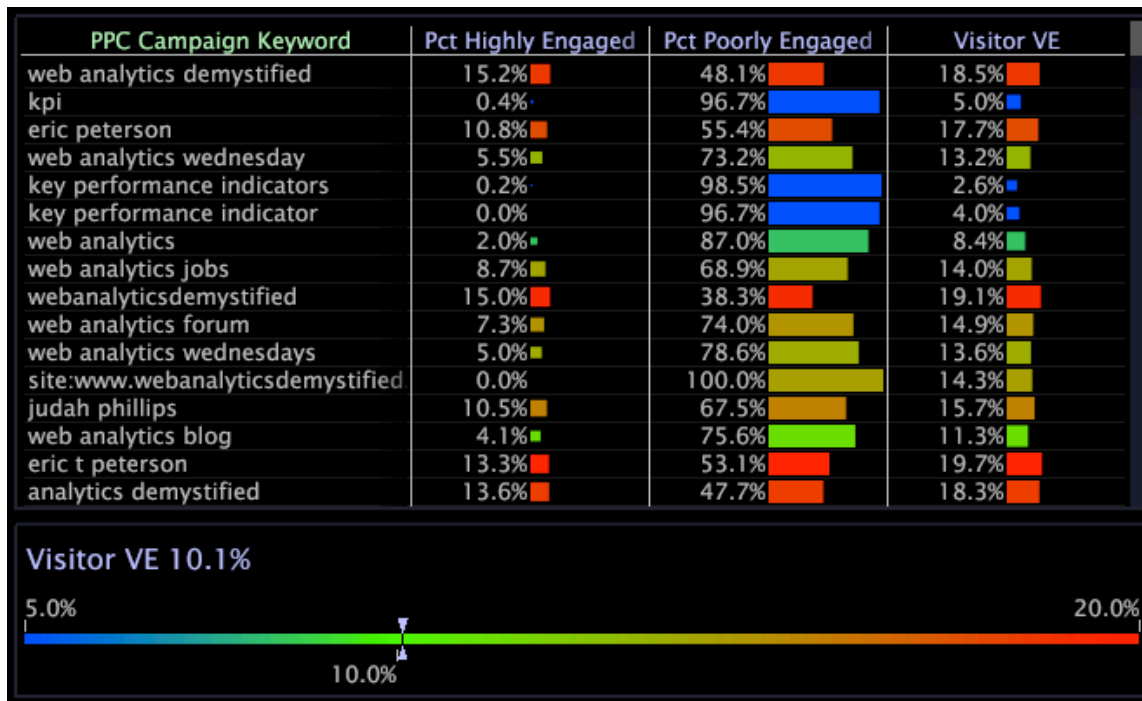


Figure 18: Example of the Visitor Engagement calculation applied to search engine keyword campaigns, showing the percentage of “highly” and “poorly” engaged visitors coming from each campaign keyword.

Marketers also benefit when using the visitor engagement metric to evaluate landing pages and subsequent content designed to draw visitors further into the site, regardless of the ultimate outcome (Figure 19).

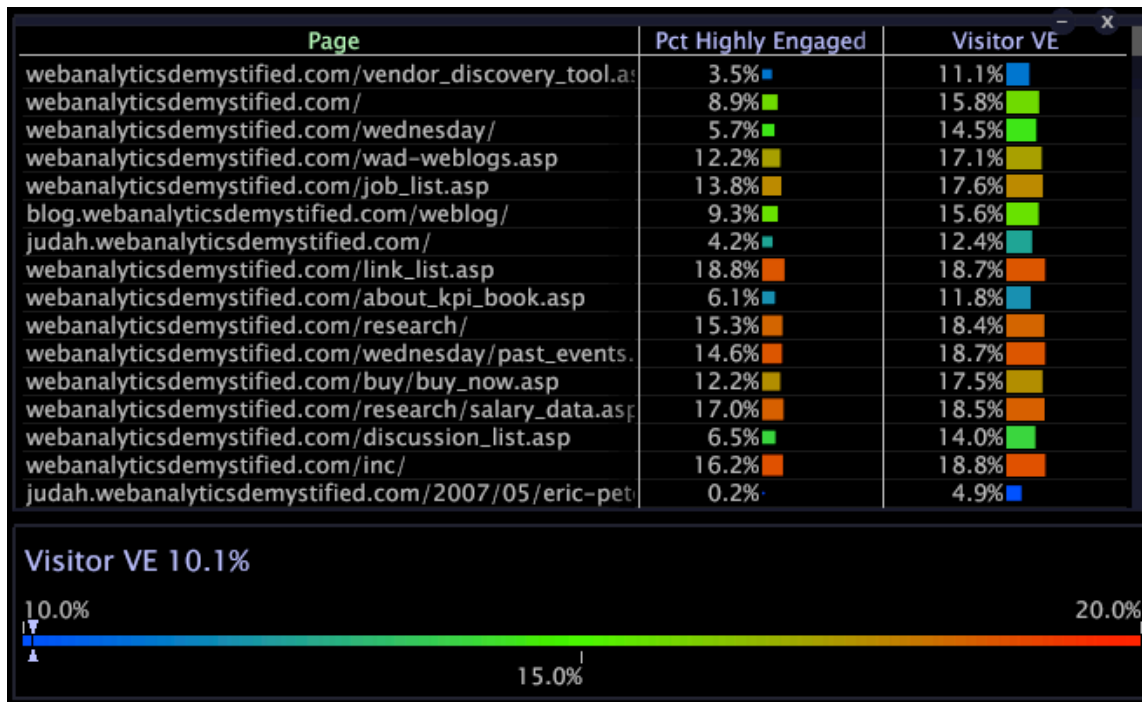


Figure 19: Example of Visitor Engagement applied to the page dimension, showing differential levels of engagement on a page-by-page basis.

Marketing often looks for new ways to assess the level of opportunity in different markets, something especially difficult when using traditional measures of success given that transactions may not be enabled for entire regions of the world. Given sophisticated enough software, the Visitor Engagement metric can be applied to this problem as well (Figure 20).



Figure 20: Example showing Visitor Engagement mapped onto a three-dimensional representation of the Earth. The areas where the circles are brighter indicate a higher level of engagement in that particular region based on the calculation described in this document (socio-political, socio-economic, and population dynamics influences are determined to be “0” in this image as it is a snapshot, not a trend image.)

Intelligent advertisers are also keeping track of what visitors are doing after responding to their ads, oftentimes by tracking their landing and entry pages. Visitor Engagement provides an entirely new lens through which advertisers can view the traffic they’re acquiring (Figure 21).

Page	Entry Rate	Pct Highly Engaged
judah.webanalyticsdemystified.com/category/appe	72.4%	0.3%
blog.webanalyticsdemystified.com/weblog/2008/0	59.6%	1.3%
judah.webanalyticsdemystified.com/2007/05/no-s	49.4%	0.8%
blog.webanalyticsdemystified.com/weblog/2007/0	46.9%	2.8%
blog.webanalyticsdemystified.com/weblog/2007/0	44.8%	2.9%
judah.webanalyticsdemystified.com/2007/05/eric-	42.1%	0.2%
blog.webanalyticsdemystified.com/weblog/2006/1	39.8%	5.0%
blog.webanalyticsdemystified.com/	39.5%	5.4%
blog.webanalyticsdemystified.com/weblog/2007/0	38.2%	6.2%
webanalyticsdemystified.com/discussion_list.asp	37.4%	6.5%
blog.webanalyticsdemystified.com/weblog/2007/1	36.6%	8.1%
webanalyticsdemystified.com/about_kpi_book.asp	34.2%	6.1%
webanalyticsdemystified.com/	32.7%	8.9%
webanalyticsdemystified.com/wednesday/	27.7%	5.7%
judah.webanalyticsdemystified.com/	26.2%	4.2%

Figure 21: Example showing entry rate and the percentage of highly engaged visitors for each of the top entry pages to a site.

Additionally, advertisers will likely benefit from being able to quickly assess the quality of traffic coming from the various outlets where they are running ads (Figure 22).

Referrer	Pct Sessions	Pct Poorly Engaged
google.com	22.6%	86.7%
live.com	11.3%	99.3%
yahoo.com	2.8%	83.4%
google.co.uk	2.3%	91.1%
stumbleupon.com	1.6%	99.7%
webanalyticsassociation.org	1.1%	72.9%
google.ca	0.8%	85.2%
google.co.in	0.8%	87.7%
wikipedia.org	0.7%	85.4%
blogspot.com	0.6%	77.1%
google.de	0.5%	86.3%
webanalytics.ox2.eu	0.5%	93.4%
netvibes.com	0.5%	85.1%
google.com.au	0.5%	91.0%
msn.com	0.4%	97.0%

Figure 22: Example showing traffic distribution for referring domains and the percentage of referred visitors categorized as “poorly” engaged using the Visitor Engagement calculation. In this example, the Web Analytics Association is the best source of traffic and StumbleUpon is the worst.

Business owners are constantly faced with making decisions about which content and functionality to keep, which to cut, and which of the seemingly unending barrage of Web 2.0 technologies to deploy. While traditional measures of interest such as page views, bounce rate, and duration appear on the surface to be good indicators of the popularity of content and technology, these metrics can be misleading. In situations where content is rendered in new ways, the measure of Visitor Engagement is likely to be more robust and more useful overall than any traditional measure that may not account for event-driven interfaces and interactions with syndicated (off site) content (Figure 23).

Page	Pct Highly Engaged	Page Views per Session	Page View Duration
webanalyti	3.5%	4.13	06:34
webanalyti	8.9%	2.25	04:42
webanalyti	5.7%	2.56	08:58
webanalyti	12.2%	2.24	04:12
webanalyti	13.8%	2.85	03:30
blog.weba	9.3%	2.07	09:34
judah.web	4.2%	2.25	06:04
webanalyti	18.8%	3.00	05:32
webanalyti	6.1%	2.12	04:18
webanalyti	15.3%	2.31	02:57
webanalyti	14.6%	31.28	00:17
webanalyti	12.2%	2.26	03:05
webanalyti	17.0%	4.48	02:28
webanalyti	6.5%	1.85	06:58
webanalyti	16.2%	1.89	02:06
judah.web	0.2%	2.21	01:47

Figure 23: Example showing differences observed when using Visitor Engagement, page views per session, and page view duration to evaluate individual pages and applications on a site.

Furthermore, business owners with indirect monetization models (content sites, marketing sites, customer support sites) and sites not directly targeting consumers (for example, B2B sites like IBM, SAP, and Oracle) will benefit significantly from the Visitor Engagement calculation which provides a much more robust measure of opportunity than page views, session duration, or any other single proxy for conversion (Figure 24).

Domain	Visitor VE	Pct Highly Engaged
tcsu.org.uk	21.6%	17.9%
level3.net	17.1%	9.6%
satama.com	15.7%	8.7%
fastwebnet.it	14.6%	7.7%
hitbox.com	17.0%	6.4%
comcastbusiness.net	13.2%	5.8%
towerstream.com	14.7%	5.5%
rcn.com	12.3%	5.5%
comcast.net	12.1%	5.4%
exacttarget.com	18.8%	5.0%
twtelecom.net	12.6%	5.0%
swbell.net	11.2%	4.9%
mindspring.com	12.4%	4.7%
webtrends.com	14.3%	4.7%
verizon.net	10.6%	4.6%

Figure 24: Example showing Visitor Engagement mapped against known visitor domains looked up using visitor IP address.

Depending on the specific business and engagement model, measures of engagement can be taken right to the level of the individual visitor, providing a previously unheard of level of granularity where opportunities exist to directly connect with prospects (Figure 25).

Visitor Email	Visitor VE	Session VE	Momentum
matt lillig	15.2%	43.5%	28.2%
coreymatthew	11.9%	18.7%	6.8%
e.hummels@	18.6%	25.3%	6.7%
cox@userint	7.7%	12.6%	4.9%
chjsmi@gma	36.9%	40.2%	3.4%
ganidjar@un	26.9%	29.7%	2.8%
gregory.bria	33.7%	35.4%	1.7%
bronkin@cor	3.4%	4.7%	1.3%
ehansen@sit	17.3%	18.6%	1.3%
szakrajsek@	14.8%	15.9%	1.0%
julien.coquet	29.2%	30.2%	0.9%
mspark@iqv	30.5%	30.6%	0.1%
wdaveonline	0.0%	0.0%	0.0%
drogers01@g	0.0%	0.0%	0.0%
mSPAETH@CO	12.9%	12.9%	0.0%

Figure 25: Visitor Engagement mapped to the level of the individual, which may or may not be appropriate depending on your business model. Also shown here is the concept of “lifetime” (= Visitor) and “session” engagement, allowing the calculation of something called “Engagement Momentum” to highlight how an individual’s level of engagement changes over time.

VISITOR ENGAGEMENT AND SEGMENTATION STRATEGIES

The most fundamentally important use of the Visitor Engagement metric is as a basis for visitor segmentation. This segmentation, and the ability to differentiate referring sources, campaigns, keywords, and pages by the measured level of visitor engagement opens the door to whole new opportunities for marketing, advertising, and site design. Similar in many ways to Recency, Frequency, and Monetary Value (RFM)⁽¹³⁾ the Visitor Engagement calculation provides the basis for exploring your audience along multiple dimensions (Figure 26).

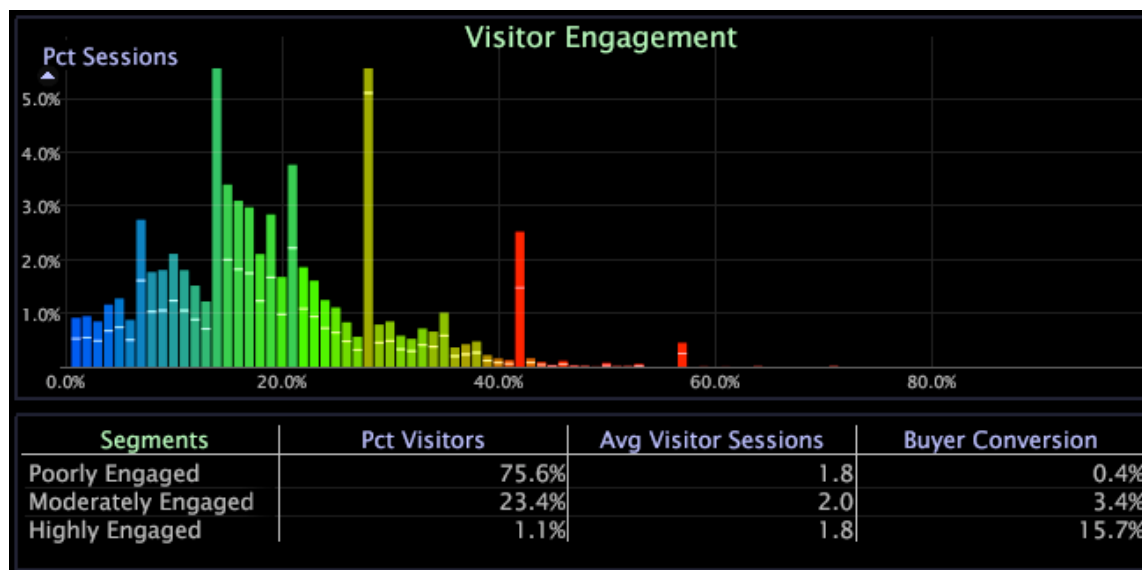


Figure 26: Example of simple Visitor Engagement segments, delineated somewhat arbitrarily at the 20% and 40% levels of measured engagement. These segments can now be applied to any other dimension in the system and combined with other segments. Interestingly, this segmentation highlights the relationship between visitor engagement and buyer conversion showing that highly engaged visitors are nearly 39 times as likely to purchase as poorly engaged visitors and 3.6 times as likely to purchase as moderately engaged visitors.

Because the Visitor Engagement calculation includes both recency and frequency (essentially the Loyalty Index applied to a fixed timeframe) but *excludes* monetary value, the measure lends itself to a less commerce-centric view of the opportunity afforded web sites. For example, Figure 27 shows a segment of “Highly Engaged” visitors from the USA, Great Britain, and Canada and lists their email domains, a potentially very valuable segment as a global sales organization reaches out looking for the best opportunities.

¹³ <http://en.wikipedia.org/wiki/RFM>

Visitor Email Domain	Visitors	Visitor VE
digiskey.com	1	71.4%
redeye.com	1	64.3%
earth.care2.com	1	57.1%
dix.me.uk	1	57.1%
ovrdrv.com	1	57.1%
circlebuilder.com	1	57.1%
mccann.com	1	57.1%

Country	Pct Visitors	Pct Sessions	Buyer Conversion
usa	66.5%	75.6%	13.8%
gbr	5.4%	3.5%	16.7%
can	4.9%	3.6%	18.5%
nld	3.1%	2.0%	29.4%
ind	2.3%	2.0%	0.0%
deu	2.3%	1.5%	7.7%

Segments	Pct Visitors	Avg Visitor Sessions	Buyer Conversion
Poorly Engaged	32.8%	1.8	0.4%
Moderately Engaged	10.7%	2.0	3.3%
Highly Engaged	0.5%	1.9	14.3%

Figure 27: Visitor segment created by selecting “Highly Engaged” visitors and visitors from the USA, Great Britain, and Canada and then applying that segmentation to the visitor’s email domain sorted by the domain’s level of Visitor Engagement.

Clearly, the segments that are created depend heavily on the type of data available in the system, as does the use of those segments. At Web Analytics Demystified, the Visitor Engagement calculation and “highly engaged” segment is used to generate an ongoing list of prospect email addresses and domains. This segmentation allows sorting through hundreds of email addresses and thousands of visitors coming to www.webanalyticsdemystified.com and identifying those individuals and companies that are most likely to inquire about consulting services.

ABOUT THE AUTHORS

For more information about this paper or the work contained herein, please contact Mr. Carrabis (jcarrabis@nextstageevolution.com) or Mr. Peterson (eric.peterson@webanalyticsdemystified.com) directly.

JOSEPH CARRABIS

Joseph Carrabis has authored 25 books and over 400 articles in five areas of expertise. His books have covered cultural anthropology, database technology and methods, information mechanics, language acquisition, learning and education theory, mathematics, network topologies, and psycholinguistic modeling. His articles have covered computer technology, cultural-knowledge modeling, equine management, knowledge studies and applications, library science, martial arts, myth and folklore, neurolinguistic, psychodynamic and psychosocial modeling, studies of group and tribal behavior, studies of social interactions in NYC and more. His knowledge and data designs have been used by Caltech, Citibank, DOD, IBM, NASA, Owens-Corning and Smith-Barney among others.

Mr. Carrabis has been everything from butcher to truck driver to Senior Knowledge Architect to Chief Research Scientist. Currently he is Chairman and Chief Research Officer of NextStage Evolution, LLC, NextStage Global LTD, and a founder of KnowledgeNH, NH Business Development Network and the Center for Semantic Excellence. He was selected as a [Senior Research Fellow and Advisory Board Member](#) to the Society for New Communications Research, is a Founding Member, Director, Predictive Analytics and Research Fellow of the [Center for Semantic Excellence](#) and has been accepted in the [Scientists Without Borders](#) program of the New York Academy of Science and the UN Millenium Project. Carrabis is the inventor and developer of Evolution Technology.

ERIC T. PETERSON

Eric T. Peterson has worked in the field of web analytics since 1998 and since that time has been employed by some of the best known companies in the industry including WebTrends, WebSideStory, and Visual Sciences. Mr. Peterson has also worked as a senior analyst at JupiterResearch covering a variety of technologies including web analytics, site search, and content management systems. He is the author of three books, *Web Analytics Demystified*, *Web Site Measurement Hacks*, and *The Big Book of Key Performance Indicators*. Peterson is a frequent speaker on the subject of measurement and measurement systems at events like Emetrics, Internet Retailer, and Shop.ORG, and has delivered hundreds of presentations on the subject around the globe.

ABOUT NEXTSTAGE GROUP

The NextStage Group of Companies provide advanced behavioral analysis solutions based on over 20 years of research and development. NextStage Evolution, based in Nashua, NH, and Scotsburn, NS, is the research division of NextStage and focuses on how people interact with digital media. NextStage Global, based in Scotsburn, NS, is the commercial division of NextStage. NextStage Global provides commercial products, custom software development, training and consulting services based on

research conducted by NextStage Evolution. Their solutions have been used to address business problems in marketing, education and security.

ABOUT WEB ANALYTICS DEMYSTIFIED

Web Analytics Demystified, founded in 2007 by internationally known author and former JupiterResearch analyst Eric T. Peterson, provides objective strategic guidance to companies striving to realize the full potential of their investment in web analytics. By bridging the gap between measurement technology and business strategy, Web Analytics Demystified has provided guidance to hundreds of companies around the world, including many of the best known retailers, financial services institutions, and media properties on the Internet.

APPENDIX

Mr. Joseph Carrabis provided the following information after an independent validation of Mr. Eric Peterson's Visitor Engagement calculation in an effort to validate the underlying mathematics in the equation and expand upon the measure's utility across a wide range of systems. Questions about the following content are best addressed to Mr. Carrabis (jcarrabis@nextstageevolution.com)

First and most importantly, my goal in working with Mr. Peterson was to take his calculation and increase its flexibility and applicability. Flexibility and applicability – basically allowing the calculation to provide useful results in a greater number of analytic situations – means making the calculation a) more rigorous, b) extendable and c) extensible mathematically. Readers comfortable with “flexibility and applicability” can skip this section and proceed to “Rule Set”. Readers curious about rigor, extendibility and extensibility may find this section helpful.

Rigorous means I made some changes to the Visitor Engagement calculation such that when it is used it provides a consistently defined result set. Much of adding *rigor* simply comes down to making sure units such as time, visitor, page counts and views, etc., are consistently defined. Using “weeks” as a time measure for one variable and “hours” for another, then adding the two variables will produce a result, yes, and that result will vary so wildly from calculation to calculation that adding this kind of *rigor* becomes mandatory. This is often called “Conservation of Units” in mathematical physics texts.

Extendable means I modified the original form so that it could accommodate special cases. Special cases might occur when a client company wants to incorporate a variable into the Visitor Engagement calculation based on their specific business needs or requirements. Creating a framework that allows for extensibility insures that Mr. Peterson's calculation will have “long legs” and not be limited to only certain tools or methodologies.

Extensibility first means the redefinition of the calculation allows for changes in the variables used if the variable set is either undesirable or unavailable. Like all flexible things, there are some caveats to how flexible things can be. Here the flexibility in variable substitution means that some basic properties must be shared between the original and substituted variables. For example, session duration can't be directly substituted for total pages viewed because the former is a time scalar and the latter is a page scalar; they are incompatible metrics. However, if we know that the average visitor traverses nine pages during their session we can substitute the former for the latter because

$$\text{Total Pages-Viewed} = \{(9 \text{ pages/visitor}) * (\text{Total Time-On-Site})\} / (\text{Time-On-Site-Per-Visitor})$$

The above is necessarily simplistic to demonstrate the first aspect of extensibility. Note that it also demonstrates the Conservation of Units principle.

Extensibility also means making Mr. Peterson's calculation able to provide meaningful results regardless of platform being measured. This was done because the web is changing rapidly and one of the ways it is changing is by allowing people to navigate web interfaces on non-traditional browsers (iPhones, SmartPhones, PDAs, etc.). Providing extensibility (in this sense) means Web Analytics Demystified's Visitor Engagement calculation (and all derivations) will be valid on a variety of devices so long as the variable definitions being used are valid on those devices.

Finally, Peterson's original formulation only allowed for discrete variable states. The modifications in this appendix allow for continuously variable states making real-time, rolling, "sum of least averages" and similar calculations possible should the need arise.

RULE SET

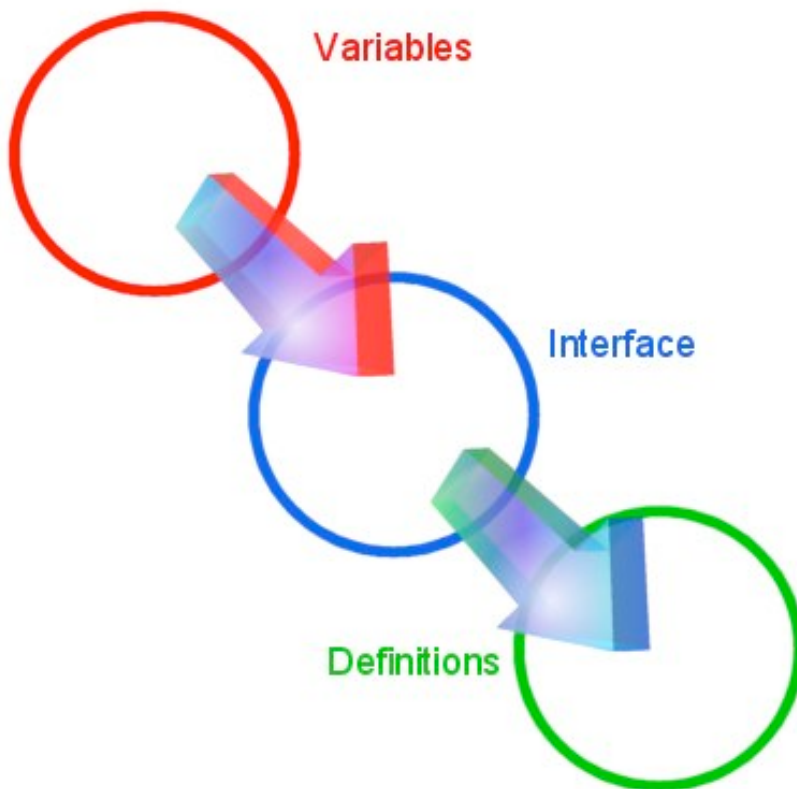
We come to defining an initial rule set that will allow for this flexibility. Here I synthesized a tripartite rule set based on Mr. Peterson's original writings on the subject on "online engagement" to the following:

Can you define what you mean?

Can you repeatedly measure what you mean, get values that are explainable by your definition and demonstrate only statistically allowable variance?

Can you make these measurements through any interface?

This accomplished and understood, the next step is to recognize that Peterson's original definition and resulting calculation define a mapping function in that he uses variables that are specific to a given interface that can be used to define "engagement" for that interface. It looks like the following image and can also be stated as "There are variables that only map to specific interfaces, therefore certain interfaces can only have specific definitions of 'engagement'."



Mr. Peterson's definition of "engagement", "*Engagement is an estimate of the degree and depth of visitor interaction on the site against a clearly defined set of goals*" is essentially the statement "There are variables that are specific to a given interface that can be used to define 'engagement' for that interface."

Peterson basically created a 1-1-1 relationship (as in "variables-interface-definition") because he took as initial conditions specific variables and a specific interface (a website). Very good if all you care about are the interfaces we recognize today, not as good if you want to have definitions that will work on any interfaces coming along through time.

The next part is to make sure the variables map to the current interface and are sufficient for the definition of "engagement" being used for that interface while obeying the basic rule set required for rigor, extendibility and extensibility.

FIRST MODIFICATION: STANDARDIZE TIME

The first modification we make sets the precedent for all modifications that follow; we take the C_i definition and require it to adhere to a very specific quantity of time. It is rewritten as:

C_i == is the percent of sessions having more than "n" page views divided by all sessions for some given period of time, t, divided by all sessions regardless of number of page views for that same period of time, t.

This means C_i now looks like this:

$$C_i = \left. \frac{\sum_{i=1}^N m_i}{m_n} \right| \begin{array}{l} m_i \in \{0, 1\} \\ \Rightarrow P_i/S_i \geq n \Rightarrow m_i = 1 \\ P_i/S_i < n \Rightarrow m_i = 0 \\ m_n \equiv \text{Total \# of sessions} \end{array}$$

The $|_t$ in the above formula merely means that all variables in the formula must use the same period of time, t, for the calculation. This modification allows you to remove the time consideration from the calculation completely once the variables are collected and rationalized to some consistent time period, t.

"m" is used as the variable in the formula because we are counting the number of sessions that match a threshold criteria – the number of page-views per session above some value, n, and I didn't want to cause counter confusion. This is also why the summations are done from 1 to N.

The above rewrite indicates that $m = f(n)$ such that $m \in \{0, 1\}$ for any value of n. Further, n is itself based on two values, page-views and sessions, thus $n = g(P_i, S_i)$ and this is where the constraint on time, t, first makes itself known. There must be a defined time interval, t, in which we are determining "all sessions, S" takes place and that each S_i also falls within that same time period.

SECOND AND THIRD MODIFICATIONS: GROUP LIKE VARIABLES TOGETHER, INCORPORATE SCALING

I digress from Mr. Peterson's original variable ordering slightly and only to group like variables together for easier calculation. "Like variables" are those with similar constraints (time, t , and session, m_n). "Scaling" is a recognition that some variables reveal much more information if they become the sums of separately valued sub-variables rather than either single valued or binary valued variables.

The F_i definition is valid so long as the metric only deals with whether or not a visitor provided feedback *and doesn't take into account the methodology of that feedback*. Under that condition it has a form similar to G above:

$$F_i = \left(\begin{array}{c} \sum_1^N f_i \\ m_n \end{array} \right)_t \quad \left. \begin{array}{l} f_i \in \{0, 1\} \\ \rightarrow F \rightarrow f_i = 1 \\ \sim F \rightarrow f_i = 0 \end{array} \right\}$$

This initial condition also allows us to use several elements from the G calculation above; the " m_n ", the standardization of f_i and indexing from 1 to N simplify the final calculation immensely. However, users are warned that this simplification (like all simplifications) loses a great deal of information that is potentially valuable to the client. This formula becomes much richer and more robust with complimentary frameworks in place.

Note that what follows borrows from NextStage research and doesn't require knowledge of NextStage research methodologies or analytics to be useful to the reader of this paper.

The simplest complimentary framework is to recognize that different response methodologies require different levels and types of conscious activity on the part of the visitor. For example, a visitor who writes a lengthy email is more psycho-cognitively engaged than someone who is given a response form with two buttons – "I liked the site", "I didn't like the site" – clicks one and submits the form. An example bifurcation of the Feedback index can be that the visitor used their own language to provide feedback, i.e., they emailed, called, filled in a form text field or performed some other similar activity that demonstrated a desire to be understood rather than a willingness to be pigeon-holed.

This simple bifurcation reconstructs the F variable to:

$$F_i = F_e + F_{\sim e}$$

Some readers might stop here and declare that I'm using "engagement" to define "engagement" and I invite them to look more closely. The definition of *engagement* used here and identified by the subscripts "e" and " $\sim e$ " is not Peterson's definition of engagement.

It is worth looking at this bifurcation closely because it provides many paths for exploration. First, the bifurcation can be further bifurcated into as many different F variables as the client's system can reasonably support; a system that supports ten different feedback mechanisms can have ten F variables, one for each feedback mechanism (examples of this concept are demonstrated for the I and L variables below).

These different F variables can be assigned different multipliers ($F_{e10} = 10 * F_{e1}$, for example) to demonstrate different visitors' greater values to a client.

Second, a client can determine which feedback mechanism is providing greater value simply by identifying each mechanism and matching that back to the value derived from all visitors using that mechanism.

Third, clients will now have the capability of isolating all other factors by various F factors; for example, do people providing feedback via F_{e5} interact with a site differently than visitors providing feedback via F_{e1} ?

Similarly, we will transform the Brand Index (B_i) using the methodology demonstrated with the F variable above; we bifurcate it into recognizable and easily identifiable components:

$$B = B_e + B_{-e}$$

Like F above, B appears in and can be used in Peterson's original calculation (with some necessary modifications) as:

$$B_i = \frac{\sum_1^N b_i}{n_n}$$

$b_i = \{\emptyset, 1\}$
 $\rightarrow B \rightarrow b_i = 1$
 $\sim B \rightarrow b_i = \emptyset$

Note that the above can be further expanded (as is demonstrated in the **Final Form** section of the Appendix) to incorporate any number of "brands."

Regarding the Interaction Index (I_i) the greatest concern with the existing definition of this is that by separating "I" type events from "conversions" we create a discontinuity in the sales cycle. Most people's goal for engagement is to get people so "engaged" they convert and if "I" type events are distinct from conversion events then we don't learn the tipping point, per se, and the ability to isolate the one (or more) elements that need to be adjusted in order to increase conversions is lost. I point to *Usability Studies 101: Defining Visitor Action* ⁽¹⁴⁾ as suggested reading for this.

Again, I believe there's a simple solution for this; a step scale for events required for a "conversion". For example:

Event Designation	Measurable Event	Value
A	Landing Page	0

¹⁴ <http://www.imediaconnection.com/content/6330.asp>

B	Navigates site	1
C	A+B	2
D	Navigates Product Path	3
E	A+D	4
F	Requests Email Follow Up	5
G	C+F	6
H	E+F	7

(Note that this step scale concept can also be applied to the F and B variables above)

What this becomes (progressing from what I've demonstrated here) is a sales funnel⁽¹⁵⁾. Mr. Peterson writes, "... is designed to provide information about the large number of visitors who don't convert" which is a valid goal. Remembering that the original summation is over all visitors, this methodology provides a clue as to where the great unwashed are in the sales cycle, i.e., how "engaged" the majority of visitors are. Now you have a basis to determine where to start your (for example) your A/B testing without burning through your research dollars on a whole site or entire page. Another, more obvious benefit from the above concept is that it allows Peterson's I_i metric to comply with our conservation of units principles from before:

$$I_i = \frac{\sum_i^N i_i}{m_n t}$$

Note that this rewrite subsumes Mr. Peterson's original S_i metric which was shown to be a similarly measurable interaction and not justifiably distinct.

The L_i metric's table is a variation on the Interaction Index, essentially a straight count of visitor sessions as shown below

# of individual visitor's sessions	Value
1	0
2	1
3	2
4	3
5	4

¹⁵ <http://www.imediaconnection.com/content/6252.asp>

6

5

7

6

The beauty in this rethinking of L_i is that you can use the above to create version A and, with a little modification, version B:

$$\text{Version A} \\ L_i = \int_t \frac{\sum_1^N l_i}{m_n}$$

$$\text{Version B} \\ L_i = \int_t \frac{\sum_1^N l_i}{L} \quad | \quad L \equiv \text{Avg \# of visits to conversion}$$

Version A is cleaner because it continues with the simplification of the final equation although and as noted previously, the simplification removes information that the client might find useful.

Version B offers the ability to recognize the “tipping point” from loyalty to disloyalty, if you will, or for the purposes of the Visitor Engagement formula when “engagement” becomes “disengagement” because “L” is the average number of visits for all visitors within the given time frame, t , that participate in a “conversion” event. Note that the event doesn’t have to be “conversion”; it can be any event or group of events you want to query. For example, if you have 50 conversions in some time period and the average number of visits per converted visitor is 5 but your average number of visits per visitor is 10 and you’ve had 500 visitor sessions during that period? It’s probably time to investigate those non-converting visitors. They’re doing something and despite all other metrics, it’s something to pay attention to.

FOURTH MODIFICATION: NON-STANDARDIZABLE VARIABLE FORMS

Mr. Peterson’s R_i and D_i variables don’t lend themselves to the simplifications demonstrated in Second and Third Modifications because they’re not session (m_n) based. Their construction is very similar none-the-less.

R_i becomes:

$$R_i = \int_t \frac{c_i}{\Delta T_i} \quad | \quad f(\tau_i) \leq f(t)$$

A concern in the definition of R_i is that it allows for some division-by-0 blowups. This is alleviated by separating the time period being investigated, t , from the time since last session, ΔT_1 . The above makes a few changes to Peterson's definition. The time, T_1 , is always less than or equal to the time period being investigated, t .

In other words, if you're investigating visitor "engagement" over a day's period, $t == \text{"day"}$ and $T == \text{"hours"}$, if $t == \text{"hours"}$ then $T == \text{"minutes"}$, etc. Note that T must be less than t : if $t = 1$ hour then T must be less than 60 minutes, etc.

Also, Mr. Peterson has "... (summed over all visitor sessions in the timeframe under examination)." This is actually handled in the summation over all visitors and doesn't need to be repeated here hence is not included in the formula above.

D_i combines some of the previous logics in that a similar time constraint, T , is introduced to accomplish Peterson's stated objectives. Note that a significant difference is that the reformulated D_i uses " d_z " in the denominator as this particular metric suffers from the stated concern of being a recounted C_i metric otherwise. Conservation of units is retained because d_z is a subset of m_n :

$$D_i = \left| \frac{\sum_1^z d_i}{d_z} \right| \begin{array}{l} d_i \equiv \{0, 1\} \\ \Rightarrow \Delta T_2 \geq \alpha \Rightarrow d_i = 1 \\ \Delta T_2 < \alpha \Rightarrow d_i = 0 \\ d_z \equiv \text{Total \# of measurable sessions} \end{array} \quad \left| \begin{array}{l} f(T_2) = f(T_1) \end{array} \right.$$

Again, I'm suggesting a minor modification to the original definition to conserve units. The modification is that d_z be all sessions with a measurable duration both within the given timeframe and of similar time unit to that used in R_i rather than simply all sessions.

This necessary modification, like all others, can be ignored at the loss of information. This reformulation does allow for scaling methodologies similar to those demonstrated previously in this document along with the benefits such methodologies incur.

FINAL FORMS

A final form (and strictly for the purposes of this whitepaper) of Web Analytics Demystified's Visitor Engagement calculation is:

$$E = \left| \frac{\sum_{j=1}^v w_j (1 + 1/\Delta T_j) + \left(\frac{\sum_l L_l}{f} \right) + \left(\frac{\sum_f F_f}{f} \right) + \left(\frac{\sum_i I_i}{i} \right) + \left(\frac{\sum_b B_b}{b} \right)}{m_v} + \sum_{k=1}^z \frac{d_k}{d_z} \right|$$

The first thing to note is that this form, although richer than the original, does resolve to Mr. Peterson's original equation when L_i , F_i , I_i and B_i become either single-valued or binary-valued variables. **Clients currently using Web Analytics Demystified's original form and deriving adequate value need not trouble themselves.** However, as their needs grow, as their requirements evolve, and their ability to collect and process data increases, the above form provides several layers of richness that would otherwise be unavailable to them.

Second, the above form allows clients to explore Branding, Loyalty, Feedback and Interactions with a great deal more subtlety, almost surgically, as their markets shift and move.

Third, this form easily grows into the more advanced form (not included in this paper) when Web Analytics Demystified clients are ready to use this metric on different platforms, interfaces and related multi-channel environments. Companies interested in exploring the use of the Visitor Engagement calculation across different platforms and interfaces are encouraged to contact Mr. Peterson and/or Mr. Carrabis directly.

Fourth, this form also allows Mr. Peterson to easily incorporate other consultants' "engagement" metrics into his calculation as requirements need, and allows him to translate his results to other tool and consultant results when moving between clients. The only requirement to this element is that only those "engagement" metrics obeying the initial three rules outlined in this appendix can be applied.

In the main text of this document suggestions are made that allow for even greater flexibility, specifically taking the concept of Branding (et al) and recognizing that each separate method of brand interaction demonstrates a pre-conceived concept of the brand itself, hence demonstrate pre-existing levels of "engagement". These separate and unique concepts of Brand, Loyalty, Feedback, etc., directly affect all remaining elements such that it would also be a valuable reformulation to use the following as a base equation:

$$E = \frac{1}{t} \left\{ \left(\sum_{j=1}^v \frac{m_j (1 + \%DT_j)}{m_v} + \sum_{k=1}^z \frac{dk}{dz} \right) + \sum_p \left[\begin{array}{c} \text{"} \\ \text{] } \end{array} \right] + \sum_f \left[\begin{array}{c} \text{"} \\ \text{] } \end{array} \right] + \sum_i \left[\begin{array}{c} \text{"} \\ \text{] } \end{array} \right] + \sum_b \left[\begin{array}{c} \text{"} \\ \text{] } \end{array} \right] \right\}$$

Here the value in each of the square brackets (“[]”) is the base formula in the first line in parentheses (“()”). Each summation of L (for example) holds some predetermined Loyalty act (or “demonstration of loyalty”) fixed. This reformulation allows greater recognition of which Loyalty (again for example) acts have higher value to the client.

Another example might be a site that allows for three loyalty acts, L₁, L₂ and L₃. L₃ tends to be only used by high-value visitors and is known to be psychologically expensive to visitors (perhaps it takes time to fill out or something). Further, L₃ is developed to provide rich information that would prove useful in understanding the motivations of non high-value visitors. Manipulating the above for L₁ allows the client to know what (if any) non visitor-value elements are needed for non high-value visitors to perform the L₃ act.

FIGURES

Figure 1: Visitor Engagement tracked across multiple sites in a domain.....9

Figure 2: Typical session duration distribution for most sites, excluding 0 to 15 seconds which typically accounts for as much as half of traffic on a session-by-session basis..... 11

Figure 3: Typical view of page views per session compared to Visitor Engagement. Note that while page views per session declined as much as 50% between March and April, Visitor Engagement was only down roughly 20% during the same period..... 12

Figure 4: Typical graph of visits per visitor accounting for the fact that most visitors never pay enough Attention to become truly engaged..... 13

Figure 5: Example Visitor Engagement set-up showing component indices and Visitor Engagement score (Visitor VE) for individual visitors to a web site..... 16

Figure 6: Distribution of Ci values for over 5,000 visitors over 27,000 sessions..... 20

Figure 7: Sample data for the Click-Depth Index reported against the referring domain dimension..... 21

Figure 8: Comparison of the Click-Depth Index (Ci) and page views per session, reported against referring domain..... 21

Figure 9: Distribution of Di values for over 5,000 visitors over 27,000 sessions..... 22

Figure 10: Sample data for Duration Index (Di) shown with Click-Depth Index (Ci) against referring domains..... 23

Figure 11: Distribution of Ri values for over 5,000 visitors over 27,000 sessions..... 24

Figure 12: Distribution of Li values for over 5,000 visitors over 27,000 sessions. The large groups occur because of the significant number of visitors in the data set that have only visited a small number of times..... 25

Figure 13: Sample of search phrases used to find Web Analytics Demystified and their associated Bi and Visitor Engagement scores. Branded terms will always have a 100% value for Bi; non-branded terms Bi values are a function of the Visitor Engagement for visitors who have used those terms..... 26

Figure 14: Sample of search engines driving traffic to Web Analytics Demystified showing their Bi and Visitor Engagement scores.	27
Figure 15: Fi scores trended on a week-over-week basis for over 5,000 visitors over 27,000 sessions.....	28
Figure 16: Example engagement goals measured via the Interaction Index. Visitors participating in these events are scored as “paying Attention” and thus their Visitors Engagement scores are relatively high (the average VE score for this site is around 10%).....	29
Figure 17: Sample of Visitor Engagement and component metrics shown against the Referring Domain dimension.....	32
Figure 18: Example of the Visitor Engagement calculation applied to search engine keyword campaigns, showing the percentage of “highly” and “poorly” engaged visitors coming from each campaign keyword.....	34
Figure 19: Example of Visitor Engagement applied to the page dimension, showing differential levels of engagement on a page-by-page basis.....	35
Figure 20: Example showing Visitor Engagement mapped onto a three-dimensional representation of the Earth. The areas where the circles are brighter indicate a higher level of engagement in that particular region based on the calculation described in this document.....	36
Figure 21: Example showing entry rate and the percentage of highly engaged visitors for each of the top entry pages to a site.	37
Figure 22: Example showing traffic distribution for referring domains and the percentage of referred visitors categorized as “poorly” engaged using the Visitor Engagement calculation. In this example, the Web Analytics Association is the best source of traffic and StumbleUpon is the worst.....	37
Figure 23: Example showing differences observed when using Visitor Engagement, page views per session, and page view duration to evaluate individual pages and applications on a site.....	37
Figure 24: Example showing Visitor Engagement mapped against known visitor domains looked up using visitor IP address.....	38
Figure 25: Visitor Engagement mapped to the level of the individual, which may or may not be appropriate depending on your business model. Also shown here is the concept of “lifetime” (= Visitor) and “session” engagement, allowing the calculation of something called “Engagement Momentum” to highlight how an individual’s level of engagement changes over time.....	38
Figure 26: Example of simple Visitor Engagement segments, delineated somewhat arbitrarily at the 20% and 40% levels of measured engagement. These segments can now be applied to any other dimension in the system and combined with other segments. Interestingly, this segmentation highlights the relationship between visitor engagement and buyer conversion showing that highly engaged visitors are nearly 39 times as likely to purchase as poorly engaged visitors and 3.6 times as likely to purchase as moderately engaged visitors.....	39
Figure 27: Visitor segment created by selecting “Highly Engaged” visitors and visitors from the USA, Great Britain, and Canada and then applying that segmentation to the visitor’s email domain sorted by the domain’s level of Visitor Engagement.....	40

